# DOCUMENT RESUME

ED 446 148 TM 031 912

| | |
|---|---|
| AUTHOR | Blum, Robert E., Ed.; Arter, Judith A., Ed. |
| TITLE | A Handbook for Student Performance Assessment in an Era of Restructuring. |
| INSTITUTION | Association for Supervision and Curriculum Development, Alexandria, VA. |
| ISBN | ISBN-0-87120-267-0 |
| PUB DATE | 1996-00-00 |
| NOTE | 789p. |
| AVAILABLE FROM | Association for Supervision and Curriculum Development, 1703 North Beauregard Street, Alexandria, VA 22311-1714. Tel: 703-578-9600; Tel: 800-933-2723 (Toll Free). |
| PUB TYPE | Collected Works - General (020) -- Guides - Non-Classroom (055) |
| EDRS PRICE | EDRS Price MF05 Plus Postage. PC Not Available from EDRS. |
| DESCRIPTORS | Curriculum Development; *Educational Change; Elementary Secondary Education; *Performance Based Assessment; *School Restructuring; *Student Evaluation |
| IDENTIFIERS | *Reform Efforts |

ABSTRACT

This handbook is designed to help school staff members and others broaden their view of assessment and put assessment into the broader context of school restructuring while keeping the focus on students. The handbook is arranged in sections, each dealing with a separate topic. Each section contains a number of papers by different authors that represent the best thinking on the topic. Each section begins with an overview that discusses the major perspectives, ideas, issues, and concerns that relate to the section topic and describes how each paper fits into this fabric. The sections are: (1) "Setting the Stage"; (2) "Placing Student Performance Assessment within the Context of School Restructuring"; (3) "Developing Student Learning Goals for the 21st Century"; (4) "Aligning Assessment with Curriculum and Instruction"; (5) "Designing Performance Tasks"; (6) "Establishing Performance Criteria"; (7) "Using Performance Assessment Information for Improvement"; and (8) "Implementing Performance Assessment." In all, there are 98 papers in the handbook and 3 "resources": a glossary, a list of assessment information sources, and a list of contributors. Most of the articles contain references and illustrations. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 446 148

A HANDBOOK FOR

# Student Performance Assessment

## *in an Era of Restructuring*

Edited by
Robert E. Blum and Judith A. Arter

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL IN
MICROFICHE ONLY HAS BEEN GRANTED BY

R. Brandt

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

2B

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM031912

BEST COPY AVAILABLE

A HANDBOOK FOR

# Student Performance Assessment

## in an Era of Restructuring

Edited by
Robert E. Blum and Judith A. Arter

ASCD
Association for Supervision
and Curriculum Development
Alexandria, Virginia

[redacted]

Association for Supervision and Curriculum Development
1250 N. Pitt Street • Alexandria, Virginia 22314
Telephone: (703) 549-9110 or 1-800-933-2723 • Fax: (703) 299-8631

Gene R. Carter, *Executive Director*
Michelle Terry, *Assistant Executive Director, Program Development*
Ronald S. Brandt, Assistant *Executive Director*
Nancy Modrak, *Managing Editor, ASCD Books*
Julie Houtz, *Sr. Associate Editor*
Carolyn Pool, *Associate Editor*
Margaret Oosterman, *Associate Editor*
Stephanie Justen, *Assistant Editor*
John Somers, *Project Assistant*
Gary Bloom, *Manager, Design and Production Services*
Karen Monaco, *Designer*
Tracey Smith, *Print Production Coordinator*
Valerie Sprague, *Desktop Publisher*

**Copyright Notice**
Copyright © 1996 by the Association for Supervision and Curriculum Development.

Because ASCD does not hold copyright for the following chapters, they may not be reproduced in any form without the consent of the authors, their publishers, or the agents listed in the copyright notice on the first page of each chapter:

I-3, I-4, I-5, I-9, I-10, II-2, II-3, II-4, II-5, II-6, III-2, III-5, III-6, III-7, III-8, III-9, III-10, III-11, III-14, III-15, III-16, III-20, III-21, IV-2, IV-7, IV-9, IV-10, IV-13, V-6, V-8, VI-2, VI-8, VI-9, VI-10, VI-11, VI-12, VI-13, VI-14, VI-15, VII-3, VII-5, VII-6, VIII-2, VIII-3, VIII-4, VIII-5, VIII-9, VIII-10, IX-1, IX-2, IX-3, IX-4

The remaining chapters may be reproduced as follows:

**For current individual ASCD members:** You may make up to 100 copies of *one chapter copyrighted by ASCD* without permission provided that (1) duplication is for one-time academic use in a nonprofit institution and (2) copies are made available without charge to recipients.

**For universities:** Requests beyond the above specifications should be submitted to the Copyright Clearance Center, 222 Rosewood Dr., Danvers, MA 01923, USA (phone 508-750-8400, fax 508-750-4470). Copies must cite the source and include an ASCD copyright notice (e.g., copyright 1994 by the Association for Supervision and Curriculum Development).

**For school districts:** Contact Mary Riendeau at (703) 549-9110, ext. 404, for permission to make more than 100 copies of any one chapter.

Organizations that wish to reprint rather than photocopy material should contact Mary Riendeau at (703) 549-9110, ext. 404. *For permission to reproduce material copyrighted by others, please contact the owner of the copyright.*

Printed in the United States of America.

ASCD Stock No.: 196020 Price: $120 S4/96

**Library of Congress Cataloging-in-Publication Data**

A Handbook for student performance assessment in an era of restructuring / edited by Robert E. Blum and Judith A. Arter.
p. cm.
Includes bibliographical references.
ISBN 0-87120-267-0
1. Educational tests and measurements—United States—Handbooks, manuals, etc. 2. Examinations—United States—Handbooks, manuals, etc. 3. Examinations—United States—Design and construction—handbooks, manuals, etc. 4. Competency based educational tests—United States—Handbooks, manuals, etc. I. Blum, Robert E. II. Arter, Judith A. III. Association for Supervision and Curriculum Development.
LB3051.H3196 1996 96-5657
371.2'71'0972—dc 20CIP

99 98 97 96 5 4 3 2 1

# A Handbook for Student Performance Assessment in an Era of Restructuring

Acknowledgments vii
Introduction ix

**I. Setting the Stage**

1. Introduction
   *Robert E. Blum and Judith A. Arter*
2. Why We Need Better Assessments
   *Lorrie A. Shepard*
3. Why Is Assessment Changing?
   *Judith A. Arter and Linda Bond*
4. Assessment
   *David T. Conley*
5. Performance-Based Assessment in the Classroom: A Planning Framework
   *Jay McTighe and Steven Ferrara*
6. Understanding the Complexities of Setting Performance Standards
   *Robert J. Marzano*
7. Technical Quality Matters
   *Joan L. Herman*
8. Authentic Assessment in the Classroom: Principles and Practices
   *Peter Winograd and Fran Davis Perkins*
9. Three Pictures of an Exhibition: Warm, Cool, and Hard
   *Joseph P. McDonald*
10. Keeping Performance Assessment in Perspective
   *Richard J. Stiggins*

**II. Placing Student Performance Assessment Within the Context of School Restructuring**

1. Introduction
   *H. Dickson Corbett and Robert E. Blum*
2. Thinking Backwards to Move Forward
   *Dickson Corbett and Robert Blum*
3. The Deep Structure of Schooling
   *Barbara Benham Tye*
4. Arizona Uses High-Skills Tests to Drive Reform in Schools
   *Education Daily*
5. Linking Restructuring to Authentic Student Achievement
   *Fred M. Newmann*
6. Exhibitions: Facing Outward, Pointing Inward
   *Joseph P. McDonald*

**III. Developing Student Learning Goals for the 21st Century**

1. Introduction
   *Robert E. Blum*
2. Workplace Basics: The Skills Employers Want
   *Anthony P. Carnevale, Leila J. Gainer, Ann S. Meltzer*

3. Civic Education for Responsible Citizens
   *Ernest L. Boyer*
4. What Really Counts in Schools
   *Elliot W. Eisner*
5. Caring Kids: The Role of the Schools
   *Alfie Kohn*
6. Preparing for Idaho's Future
   *Paul Berman and Driek Zirinsky*
7. The New Definition of Learning: The First Step for School Reform
   *Beau Fly Jones and Todd Fennimore*
8. "Expect the Best: From Our Students, of Our Staff, for Our Community"
   *Snoqualmie Valley School District No. 410, Snoqualmie, Washington*
9. "Learning for a Lifetime": Mission and Student Goals
   *Salem-Keizer Public Schools, Salem, Oregon*
10. Shared Meaning: Student Learning Goals
   *Aurora Public Schools, Aurora, Colorado*
11. Maine's Common Core of Learning: An Investment in Maine's Future
   *Commission on Maine's Common Core of Learning*
12. National Education Goals and Objectives
   *U.S. Congress*
13. Oregon's Certificate of Initial Mastery Outcomes and Extended Outcome Definitions (Draft)
   *Oregon Department of Education*
14. Teaching a Core of Common Knowledge
   *Task Force on Education of Young Adolescents, Carnegie Council on Adolescent Development*
15. National Science Education Standards: Discussion Summary (Draft)
   *National Research Council of the National Committee on Science Education Standards and Assessment*
16. Why "Outcome" Standards Are Not Enough
   *Linda Darling-Hammond*
17. Establishing the Rudder for Change: Creating 21st Century Student Learning Goals
   *Thomas A. Olson*
18. Developing Exit Outcomes by Community Summit
   *Gary E. Dunkleberger*
19. Outcomes, Performances, and Assessment: The Agony and Ecstasy of Development
   *Terry Mazany and Pamala Noli*
20. The Discipline of Building Shared Vision
   *Peter Senge*
21. In Search of Community
   *Ernest L. Boyer*

### IV. Aligning Assessment with Curriculum and Instruction

1. Introduction
   *Judith A. Arter*
2. Reassessing Assessment: Seven Issues Facing Renaissance Schools
   *Arthur L. Costa and Bena Kallick*
3. The Instructional Resource Network: Supporting the Alignment of Curriculum, Instruction, and Assessment
   *Kathleen A. Fitzpatrick, Marilynn Kulieke, John Hillary, and Vern Begitschke*
4. Good Measure: Assessment as a Tool for Educational Reform
   *Dennie Palmer Wolf, Paul G. LeMahieu, and JoAnne Eresh*
5. When Teachers Look at Student Work
   *Christine Sobray Evans*
6. Integrating Assessment and Instruction: Continuous Monitoring
   *Jackie Palmer*
7. Linking Assessment with Instruction
   *Richard J. Shavelson and Gail P. Baxter*
8. What You Assess May *Not* Be What You Get
   *Thomas R. Guskey*
9. Evaluating the Teaching and Learning of Literacy
   *Francine Z. Stayter and Peter Johnston*
10. Using Assessment as a Tool for Learning
   *Judith A. Arter*
11. Collaborative Assessment: Making High Standards a Reality for All Students
   *Doris Sperling*
12. Retelling: Reading Assessment That's Also Good Instruction (or, Reading Instruction That's Also Good Assessment)
   *Jane B. Braunger*
13. Portfolio Reflections in Middle and Secondary School Classrooms
   *Roberta Camp*

### V. Designing Performance Assessment Tasks

1. Introduction
   *Judith A. Arter*
2. Designing an Authentic Assessment
   *Sandra Schnitzer*

3. Alternative Assessment in Science and Mathematics
   *Judith A. Arter*
4. Keys to Good Writing Prompts
   *Vicki Spandel*
5. Using Scoring Considerations to Develop Effective Performance Assessment Tasks
   *Gail Lynn Goldberg*
6. Creating Tests Worth Taking
   *Grant Wiggins*
7. What We've Learned About Assessing Hands-On Science
   *Richard J. Shavelson and Gail P. Baxter*
8. The Development of Performance Assessment Exercises
9. Validating and Anchoring Classroom Performance Tasks
   *Stephen K. Hess*

**VI. Establishing Performance Criteria**

1. Introduction
   *Judith A. Arter*
2. Performance Criteria: The Heart of the Matter
   *Judith A. Arter*
3. Empowering Students with Instructional Rubrics
   *William L. Lazzaro*
4. Setting Criteria
   *Joan Herman, Pamela R. Aschbacher, and Lynn Winters*
5. What Is a Rubric? A Dialogue on Design and Use
   *Grant Wiggins*
6. Self-Assessment Using Performance Task Assessment Lists
   *K. Michael Hibbard*
7. Writing Assessment
   *Vicki Spandel and Ruth Culham*
8. Performance Assessment in Science
   *Greg Hall*
9. Assessment of Laboratory Skills in High School Science
   *Rodney Doran*
10. Primary Developmental Continuums—Reading, Speaking, Writing
   *Annie Calkins*
11. Mathematics Problem Solving
   *Oregon Department of Education*
12. Group Performance Tasks: History-Social Science, Grade 11 Historical Investigations
   *California Learning Assessment System*
13. Thinking Skills Assessment
   *Jeanne Bond-Essner and Rachel Korach*
14. Assessing Portfolios
   *Vicki Spandel and Ruth Culham*
15. Small-Group Discussion
   *Alberta Education, Canada*

**VII. Using Performance Assessment Information for Improvement**

1. Introduction
   *Dean G. Arrasmith, James W. Kushman, and Nancey Olson*
2. Poway Portfolio Project
   *John Winbury and Christine Sobray Evans*
3. The Impact of Training Students to Be Self-Assessors of Writing
   *Judith A. Arter, Vicki Spandel, Ruth Culham, and Jim Pollard*
4. Managing Improvement by Profiling
   *Robert E. Blum and Jocelyn A. Butler*
5. Performance Assessment for Restructured Schools
   *Douglas E. Harris and Joyce Stone*
6. Assessing Student Performance in Crater High School's Applied Technology Program
   *Nancey Olson*
7. Using Standards and Assessments to Promote Excellence and Equity Among Diverse Students
   *Claude Goldenberg and Bill Saunders*
8. Moving from Seat Time to Mastery: One District's System
   *Kevin Castner, Lorraine Costella, and Stephen K. Hess*
9. Using the Results of Performance Assessment to Support School Reform and Improve Student Achievement
   *Louise F. Waynant*
10. Improving Assessment and Reporting in the Edmonton Public Schools
   *Dale Armstrong*

**VIII. Implementing Performance Assessment**

1. Introduction
   *Shirley M. Hord and Gene E. Hall*
2. Attempting to Enhance Students' Learning: A Search for Valid Programs and Reliable, Systemic Implementation Supports
   *Sam Stringfield*

3. The Central Office Role in Instructional Improvement
   *H. Dickson Corbett and Bruce Wilson*
4. A Time to Summarize: The Louisiana School Effectiveness Study
   *Sam Stringfield and Charles Teddlie*
5. The Principal's Role in Change: Technical and Symbolic Aspects of School Improvement
   *Terrence E. Deal and Kent D. Peterson*
6. What We've Learned: Reflections on Implementing Performance Assessment
   *Tim Westerberg*
7. Overcoming Barriers to Educational Change
   *Michael G. Fullan*
8. Leadership and the Process of Changing Practice
   *Shirley M. Hord*
9. Creating a Context for Change
   *Vicki Boyd*
10. Realizing School Improvement Through Understanding the Change Process
   *Shirley M. Hord*

**IX. Performance Assessment Examples**

1. Oregon's Writing Assessment
   *Oregon Department of Education*
2. Schoolwide Portfolios
   *Donald Daws*
3. Assessing What We Value: Snapshots from Waialae's Work in Progress
   *Kathleen U. Busick*
4. The Senior Project: South Medford High School
   *Nancey Olson*

**X. Resources**

1. Glossary of Assessment Terms
   *compiled by Judith A. Arter and Robert E. Blum*
2. Assessment Information Sources
   *compiled by Judith A. Arter and Jay McTighe*
3. Contributors

# Acknowledgments

A handbook of this size and scope requires the support and assistance of many people from the Northwest Regional Educational Laboratory (NWREL) and other educational agencies.

First, we acknowledge the key role of Ronald S. Brandt of ASCD in making the *Handbook* a reality. Recognizing a need for a resource of this kind, Ron initially approached us, asking if NWREL would be interested in compiling and editing such a volume. Ron also convened planning meetings, reviewed and commented on drafts, and put discussion of the *Handbook* on the agenda of meetings of the ASCD Assessment Consortium.

The ASCD Assessment Consortium members reviewed an early draft of the contents list and one of the sections; identified gaps; suggested potential topics, articles, and authors; and advised us on the format, length, and uses of the *Handbook*. In addition, Consortium members wrote several papers for the *Handbook*. We are grateful for their valuable contributions.

We extend our thanks to NWREL executive director Ethel Simon-McWilliams and former executive director Robert R. Rath for their support and commitment of NWREL resources to preparing the manuscript.

Each section of the *Handbook* came into being under the direction of a section editor or editors, who solicited, reviewed, selected, and edited the submissions, as well as wrote the section introductions. We commend the good work of these colleagues: Dick Corbett of Research for Better Schools, who coedited the section on the restructuring context; Jay McTighe of the Maryland Assessment Consortium, who coedited the section on designing tasks and contributed information on video resources; Dean Arrasmith, Jim Kushman, and Nancey Olson of NWREL, who edited the section on using performance assessment results; and Shirley Hord of the Southwest Educational Development Laboratory and Gene E. Hall of the University of Northern Colorado, who coedited the section on implementation and change. We prepared the general introduction as well as edited or co-edited the sections on current assessment issues, restructuring context, developing student goals, aligning assessment with curriculum and instruction, designing performance tasks, and establishing performance criteria. We also compiled the glossary of assessment terms and the assessment information sources.

Preparing the draft manuscript for ASCD also required a great deal of attention. We acknowledge NWREL research and development specialist Kathleen Cotton for handling communications with contributors about their submissions; compiling the list of contributors; and coordinating the review and production process, including technical editing of the entire manuscript.

We acknowledge NWREL library assistant Linda Fitch for extensive reference work, and NWREL production staff and assistants—Linda Gipe, Linda Mills, Tiffany Burton, Kathy Peterson, Matthew Whitaker, and Mike Ludeman—for preparing the manuscript and related correspondence.

Finally, and most important, we thank the more than 100 educational agencies, publishers, and assessment experts, about half of whom gave us permission to reprint articles that been published previously, and the other half of whom developed and submitted new material specifically for this handbook.

*Robert E. Blum, Director*
*School, Community,*
*and Professional*
*Development Program,*
*NWREL*

*Judith A. Arter, Manager*
*Assessment Unit*
*Evaluation and*
*Assessment Program,*
NWREL

# Introduction

ROBERT E. BLUM AND JUDITH A. ARTER
NORTHWEST REGIONAL EDUCATIONAL LABORATORY

## Why Is the Handbook Important?

This handbook is about learning—for staff and for students—and is directed at staff learning about performance assessment in the context of school restructuring. But the starting point for this topic is student learning—what, how, and how well must all of our students learn to be successful in the 21st century. The *Handbook* will help school staff members and others broaden their view of assessment and put assessment into the broader context of school restructuring while keeping the focus on students.

## Content of the Handbook

The *Handbook* is arranged in sections, each dealing with a separate topic. Each section contains a number of papers from different authors that represent the best thinking on the topic. Each section begins with an overview that discusses the major perspectives, ideas, issues, and concerns that relate to the section topic and describes how each paper fits into this fabric.

**Section I, Setting The Stage**, includes information on general assessment issues necessary to appreciate and interpret the rest of the material in the *Handbook*. The papers include:

- Definitions of assessment terms
- Reasons why educators are placing increasing emphasis on alternative assessment
- Keeping performance assessment in perspective
- Technical issues surrounding performance assessment

**Section II, Placing Student Performance Assessment Within the Context of School Restructuring**, presents a way of thinking about school restructuring. This model of school restructuring begins with the need to clarify our goals for students—what we want students to know and be able to do when they leave school. The model proceeds with designing learning and teaching experiences, assessment procedures, and school culture that align with our goals for students. This is important information because it presents the big picture of the restructuring process and the role of assessment in restructuring.

**Section III, Developing Student Learning Goals for the 21st Century**, addresses goals for students. The first step in designing assessments and in restructuring schools is to have a clear idea of what we want to accomplish with our students. What student outcomes will be enhanced through restructuring? What student outcomes will be the organizing principles for curriculum, instruction, and assessment? A great deal of activity exists in this area at all levels of the educational enterprise, from school districts to national professional organizations. The papers in this section provide information on current visions of what students should learn in various subjects.

Papers on procedures for building consensus about what students should learn are also included.

**Section IV, Aligning Assessment with Curriculum and Instruction,** presents papers on various conceptions of what it means to "integrate assessment and instruction." Educators must know the options so they can make good decisions about the nature of their assessments.

**Section V, Designing Performance Tasks,** looks closely at designing performance tasks, and both Sections V and VI provide in-depth information on developing quality performance assessments. The performance task is the activity, exercise, assignment, or problem that the student is asked to do. This section includes papers that describe various types of performance tasks, how to develop performance tasks, and how to ensure quality.

**Section VI, Establishing Performance Criteria**—Although good-quality performance tasks are an important part of the assessment, a well-defined way to evaluate the quality of student performances is essential. Section VI includes papers that explain why we need good-quality performance criteria, what performance criteria are, how to develop performance criteria, and how to ensure quality.

**Section VII, Using Performance Assessment Information for Improvement**—Assessment is an empty exercise unless used for something. Section VII presents information on the use of performance assessment information to improve learning at classroom, school, district, and state levels.

**Section VIII, Implementing Performance Assessment,** is about change processes. The move to more authentic assessment is an enormous undertaking, and we must pay attention to what research tells us about creating change in people and in organizations.

Comprehensive examples of performance assessments in the context of restructuring follow the main sections of the *Handbook*. Each example traces the progression of events, from stating a general goal for students, to developing an expanded definition of that goal, to designing performance assessments to measure that goal, to producing reports of the information obtained.

A final Resources section provides a glossary of terms; a list of agencies, state departments of education and school districts, publishers, and publications related to student performance assessment; and contact information for the contributors to this handbook.

## How to Use This Handbook

Although the *Handbook* is laid out in linear fashion, restructuring actually is nonlinear and messy. Work begun in one area expands to include other areas in a back-and-forth play. For example, some people jump on the "performance assessment bandwagon" and then realize that staff and community have differing beliefs about what is important to assess. This discontinuity may then lead to communitywide goal-setting processes. As goals become clearer, a focus on day-to-day learning experiences may emerge and shift the focus of the group to instruction for a while.

Perhaps different people and groups in the same school, district, or community will work on different aspects of restructuring at the same time. Starting points and approaches are as varied as people, schools, districts, and states. But the intent and direction of all the efforts should be to attain higher levels of student performance.

The *Handbook* is a resource for learning about school restructuring and student performance assessment. Pieces and parts should be used to promote learning among staff members and others on topics of importance at a particular time. Some schools are just beginning to restructure and the first section of this book may help to set the stage and persuade "key stakeholders" in the community to think coherently about restructuring. Other users may find excellent resources for reaching agreement on what students should learn. Still others may be working hard on developing performance assessments. The information in sections IV, V, and VI may be just what they need.

The *Handbook* can serve as a structure for organizing your own resources, too. The editors have pulled together a good starter set of papers on student performance assessment in the context of restructuring. Because new information appears daily, however, you are encouraged to add your own articles and ideas as well.

The *Handbook* is an open book. We hope that you will think of it in these terms and use it as a tool for learning.

## Use of Terms

Many assessment terms are currently in use: authentic assessment, alternative assessment, innovative assessment, benchmark, and standards, for instance. This book includes a glossary, but because various authors use these words in slightly different ways we cannot guarantee that every definition in the glossary will match the usages of all authors. Just remember that all the authors are talking about the same sort of thing—assessment that . . .

- does not typically involve multiple-choice tests;
- requires students to create more realistic responses to tasks and requires the application of skills used in combination;
- uses scoring that involves human judgments of quality more than pat right-and-wrong answers; and
- provides examples of exemplary, adequate, or inadequate work at various grade levels.

It is important to avoid creating false dichotomies. Performance assessment is not the opposite of multiple-choice tests; nor is performance assessment necessarily any more authentic than any other kind of assessment.

> There are many degrees and kinds of artificialities in tests. . . . Performance and product evaluation is one in which some criterion situation is simulated to a much greater degree than is represented by the usual paper and pencil test. . . . There is no absolute distinction between performance tests and other classes of tests—the performance test is one that is relatively realistic (Fitzpatrick and Morrison 1971).

## REFERENCE

Fitzpatrick, R., and E. Morrison. (1971). "Performance and Product Evaluation." In *Educational Measurement*, edited by R.L. Thorndike. Washington D.C.: American Council on Education.

# 1

# *Setting the Stage*

I-1

# Setting the Stage

## Introduction

ROBERT E. BLUM AND JUDITH A. ARTER
NORTHWEST REGIONAL EDUCATIONAL LABORATORY

This section includes essential reading that provides background and context for the remainder of the guide. Included are definitions of terms, the rationale for changes in assessment, designing performance assessments, technical concerns surrounding performance assessment, and putting performance assessment in perspective.

### What Is Performance Assessment?

Performance assessment is one of the hottest topics in education today. It seems as though assessment, particularly performance assessment, enters every discussion about schools—and rightly so, for assessment is the basis for determining the success of students, teachers, and schools. Developing high-quality assessments, however, has proven the most difficult aspect of improving schools.

Put simply, performance assessment is assessment based on observation and judgment (Stiggins 1994). We observe a student perform a task, or we review a student-produced product, and we judge its quality. We could observe a student's science experiment and judge the quality of the thinking involved, or we could read a student's research report in history and judge the quality of argumentation and writing.

Several of the papers in this section provide examples of performance assessments and related definitions.

### Reasons for Performance Assessment

**Lorrie A. Shepard**, **Linda Bond and Judith A. Arter**, and **David T. Conley** articulate the many reasons for the current emphasis on performance assessment. In essence, these authors say that we need to have assessments that more closely reflect the learning goals we have for students, communicate the right messages about what we value, align with current theories of instruction, and describe students instead of merely sorting them.

In the context of this guide, performance assessment is important because of the nature of the goals that many educational entities set for students. If our goals for students include such things as critical thinking, problem solving, communication, collaborative working, and lifelong learning, how are we going to assess progress? Many see performance assessment as an answer. Additionally, the development and use of performance assessments can help us sharpen the conception of our goals for students and can result in staff development for teachers. If designed carefully, performance assessments can also be used as tools for learning in the classroom (see Section IV).

## Designing Performance Assessments

Two papers, one by **Jay McTighe and Steven Ferrara** and one by **Robert Marzano**, provide excellent discussions of performance assessment design. McTighe and Ferrara focus on classroom assessment, whereas Marzano provides a framework for deciding which student outcomes are best assessed with performance assessments and discusses design considerations associated with different student outcomes.

## Concerns About Performance Assessment

Performance assessment is hardly new. Teachers base most of their assessments of student learning and progress on day-to-day observation in the classroom. What is new are the attempts to give such assessments a more central role in large-scale assessment (assessment that occurs primarily for purposes outside the classroom) and to make the judgments that occur within the classroom more systematic. The challenge in both these areas is to make subjective assessments as objective as possible; to make it more systematic and, therefore, more credible.

Many challenges—from the technical to the practical—accompany performance assessment, especially when assessments are used for large-scale, high-stakes purposes such as high school graduation. **Joan Herman** provides an excellent discussion of these challenges, illustrated by research on how current performance assessment efforts measure up. **Peter Winograd and Fran Davis Perkins** offer a similar discussion that targets primarily classroom-level assessment. **Joseph P. McDonald** provides a graphic demonstration of the positive and negative sides of a form of performance assessment called exhibitions of mastery. On the one hand, exhibitions of mastery can be celebrations of learning that leave everyone with a sense of accomplishment. On the other hand, they can lead to assessment that is not sufficiently objective to be credible.

## Keeping Performance Assessment in Perspective

Another note of caution about performance assessment is in order: Merely switching to performance assessment will not solve all our assessment problems. To an astounding number of people, the development of portfolio assessment seems to carry the same kind of symbolism as the search for the Holy Grail, that elusive something that, when we find it, will solve all of our problems (Eresh 1990). **Richard J. Stiggins** strives to keep performance assessment in perspective, maintaining that sound assessment requires that we have a clear idea of what we want to accomplish with students and then choose the best way to assess it. Appropriate assessment might sometimes be a multiple-choice test and sometimes a performance assessment.

### REFERENCES

Eresh, J.T. (1990). "Portfolio Aassessments as a Means of Self-Directed Learning." Paper presented at the American Educational Research Association Annual Conference, Boston.

Stiggins, R. J. (1994). *Student-Centered Classroom Assessment*. New York: Macmillan.

# I-2

# Why We Need Better Assessments

LORRIE A. SHEPARD

Reprinted with permission of the author from L.A. Shepard, "Why We Need Better Assessments," *Educational Leadership* 46, 7 (April 1989): 4–9.

16

LORRIE A. SHEPARD

# Why We Need Better Assessments

Educators should use a variety of assessment measures, make substantive improvements to standardized tests, and remove incentives to teach to the test.

In the United States today, standardized testing is running amok. Newspapers rank schools and districts by their test scores. Real estate agents use test scores to identify the "best" schools as selling points for expensive housing. Superintendents can be fired for low scores, and teachers can receive merit pay for high scores. Superintendents exhort principals and principals admonish teachers to raise test scores—rather than to increase learning. Occasionally school boards issue a mandate that "all students must be above the national norm," which, absurdly, is the same as requiring that 100 percent of students be above the 50th percentile.

## Limitations of Standardized Tests

Current political pressures exaggerate the inherent flaws in paper-and-pencil tests. Why should the public fear that test scores are going up without a real gain in achievement (Cannell 1987)? Because large-scale testing programs are limited by political and practical considerations that undermine their fidelity to important learning outcomes.

TEST 1
Answer Sheet
TEST 1
TEST 2
PAGE 5
64 Trade in the coupon
1) save
2) win with
3) send for
4) redeem
65 A terrible mishap
1) misunderstanding
2) mistrust
3) misfortune
4) misuse
72 Copy the letter
1) Reproduce
2) mail
3) answer
4) write
73 The national population count
1) legislation
2) census
3) tax policy
4) opinion
74 Unknown phone caller
PAGE 7
76 Which state lies directly south of Washington State?
1) California
2) Montana
3) Oregon
4) Idaho
78 The city of Washington, D.C.
Art by Simeon Montesa and Karen Bradford

BEST COPY AVAILABLE

*The negotiation of content.* First, the content of such tests must be negotiated. Most state-developed tests go through a consensus-building process: teams of curriculum experts and teachers agree to the content of the test. Publishers of standardized tests follow similar procedures; in addition, they do content analyses to ensure that test objectives are well matched to widely used textbooks. These procedures are sensible but they have a homogenizing effect, limiting both the breadth and depth of content coverage. The most imaginative and challenging problem-solving tasks offered uniquely by a single district are likely to be negotiated out of the test content domain. Ironically, the textbooks that commercial test framers look to for verification of coverage are governed by the same sort of consensus model, to ensure market appeal, and are in turn matched to the outlines of standardized tests (Tyson-Bernstein 1988).

*The narrowing of content.* Test construction is further constrained by the emphasis on basic skills, limiting the "height" as well as the depth and breadth of permissible content. Even advocates of high-stakes testing acknowledge that the tests do not cover the full range of important instructional objectives. Instead, the argument goes, "creative teachers can efficiently promote mastery of content-to-be-tested and then get on with other classroom pursuits" (Popham 1987, p. 682). The extent of this narrowing—that is, whether 10 percent or 50 percent of desirable content is sacrificed—will depend on the particular test and the accountability purpose or market that shaped it. For example, minimum competency or promotional gates tests are even narrower than standardized tests of basic skills.

*Multiple-choice formats.* Practical considerations also dictate test format in ways that further constrain content. Given the huge numbers of examinees and time limits, test developers use multiple-choice formats to the exclusion of tasks where students would produce or explain a correct answer. Essay tests of subject matter knowledge are rare. Although multiple-choice questions can elicit important conceptual distinctions from students, tests composed entirely of such items do not measure a respondent's ability to organize relevant information and present a coherent argument. Time limits and the need to survey as much content as possible also preclude problem sets aimed at the interrelation of facets of the same topic.

*Other format problems.* Similarly, reading passages are strikingly shorter and less complex than the texts students use for daily instruction. Moreover, to avoid confusing students by frequent changes in format, each test uses a small number of item types—despite evidence that students may appear to know a concept or skill when it is measured in one format but not know it if measured another way (Shepard 1988). The effects of these format limitations for the content validity of the test will increase over time, if the same test or a similar test is given each year.

**Even advocates of high-stakes testing acknowledge that standardized tests do not cover the full range of important instructional objectives.**

### Teaching to the Test

Despite these numerous drawbacks, the public gives standardized test scores great weight. When the scores have serious consequences—and they often do—teachers will teach to the test. Indeed, it is often the explicit intention of policymakers to force teachers to address essential skills. But teaching to the test cheapens instruction and undermines the authenticity of scores as measures of what children really know, because tests are imperfect proxies even for the knowledge domains nominally covered by the tests; and they also omit important learning goals beyond the boundaries of the test domain.

Researchers have found ample evidence that testing shapes instruction. In one study, for example, elementary teachers reported taking time away from science and social studies to devote more time to tested math skills (Salmon-Cox 1982, 1984). Darling-Hammond and Wise (1985) also found that tested areas were taught at the expense of untested areas; McNeil (1988) documented that minimum proficiencies dictated by competency tests overwhelmed class time.

In addition, Darling-Hammond and Wise (1985) found that even within the bounds of test-driven content there was "dumbing down" of instruction. Teachers taught the precise content of the tests rather than underlying concepts; and skills were taught in the same format as the test rather than as they would be used in the real world. For example, teachers reported giving up essay tests because they are inefficient in preparing students for multiple-choice tests.

Conceiving instruction in the format of multiple-choice items has other far-reaching negative consequences: it leads to endless drill and practice on decontextualized skills. The notion that learning comes about by the accretion of little bits is outmoded learning theory. Current models of learning based on cognitive psychology contend that learners gain understanding when they construct their own knowledge and develop their own cognitive maps of the interconnections among concepts and facts. Thus, real learning cannot be spoon-fed, one skill at a time.

Logically, then, to become adept at thinking and reasoning, students need practice in solving real problems and comprehending complex texts. Not surprisingly, students given instruction aimed at conceptual understanding do better on skills tests than students drilled on the skills directly (Carpenter et al. 1988). Thus, the practice of postponing higher-order thinking goals until low-level skills have been mastered is harmful. Low-achieving students suffer most from a proficiency-driven curriculum because they are consigned indefinitely to dull and repetitive skills instruction that does not enable them to grasp underlying concepts (Levin 1987).

Moreover, test-defined instruction has the effect of driving out good teachers and "deskilling" those who remain (McNeil 1988). Teachers with the least content knowledge might feel secure in leading students through worksheets fashioned after the tests, even if they become less skilled as teachers in the process. Teachers who know the most about content, however, are insulted when external mandates prevent them from using their own expertise to devise instruction. They must either find a way to resist the deskilling, as in McNeil's magnet schools, capitulate, or get out of teaching.

Finally, teaching to the test devalues the meaning of the test results themselves. By having students practice on "remarkably similar" items, teachers can improve their test performance, but gains from this type of teaching do not necessarily generalize to independent measures of the same content. Thus, test scores can go up without a commensurate gain in achievement. To really assess what students know, legislators and school boards would need a completely new test for each administration.

## Substantively Better Assessments

Twenty years ago, standardized tests served as reasonable indicators of student learning. In today's political climate, tests are inadequate and misleading as measures of achievement. Assessment tasks should be redesigned—indeed, are being redesigned—to more closely resemble real learning tasks. Tests should require more complex and challenging mental processes from students. They should acknowledge more than one approach or one right answer and should place more emphasis on uncoached explanations and real student products. Structured formats should be changed often enough that there can be no benefit to practicing a skill in one particular format. Further, the dimensions of the test domain must be expanded so that teaching to the test does not imply teaching only a subset of learning goals.

**Students given instruction aimed at conceptual understanding do better on skills tests than students drilled on the skills directly.**

At a recent conference on Assessment in the Service of Learning, Robert Glaser (1988) identified several indicators of competence in a domain of knowledge. Assessment should collect these types of evidence:

- *Coherence of knowledge.* Beginners' knowledge is spotty and superficial, but as learning progresses, understanding becomes integrated and structured. Thus assessment should tap the connectedness of concepts and the student's ability to access "interrelated chunks."
- *Principled problem solving.* Advanced learners ignore the surface features of a task and recognize underlying principles and patterns needed to solve the problem.
- *Knowledge use.* Complete understanding also means knowing the conditions that mediate the use of knowledge.
- *Automatized skills.* Basic component skills must be automatized so as to be integrated into total performance. (This is the only indicator that resembles today's by-rote measures of skills.)
- *Metacognitive or self-regulatory skills.* Assessment should determine whether students are able to monitor their own understanding, use strategies to make questions comprehensible, evaluate the relevance of accessible knowledge, and verify their own solutions.

The best way to check for these indicators is to make assessment measures resemble learning tasks. In current practice, some of the best examples of more learning-like tests are writing assessments. So long as language mechanics are not overemphasized and ideas and coherence of expression submerged, the test product represents important learning, and practice for the test is practice with real writing. Another way to assess achievement is by collecting portfolios of students' work. In the Vermont assessment proposal, for example, students and teachers would select samples of each student's best written work in English, social studies, and science. Besides encouraging more thoughtful activities in the classroom, such an approach can "enhance teacher professionalism by enabling classroom teachers to get involved in developing and scoring the assessment" (Rothman 1988).

### Classroom Assessment

Better content is not the only answer to improved assessment. Policymakers must also understand and preserve the important distinction between classroom assessment and accountability testing. Cole (1988) explained the incompatibilities between measurement to serve accountability and policy goals and measurement for instructional purposes. Large-scale assessments must be formal, objective, time-efficient, cost-efficient, widely applicable, and centrally processed. Most important, results must be in a form useful to policymakers, which usually means reducing complexity to a single score. In contrast, assessments designed to support instruction are informal, teacher-mandated, adapted to local context, locally scored, sensitive to short-term change in students' knowledge, and meaningful to students. They provide immediate, detailed, and complex feedback; and they incorporate tasks that have instructional value in themselves.

Classroom assessment is conducted in a climate of greater trust than are standardized tests. Classroom tests and observations do not have to meet the same standards of accuracy. Errors made in judging individual students are less serious and more easily redressed as teachers gather new evidence. Although single teacher tests are probably less reliable (in a statistical sense) than a one-hour standardized test, the accumulation of data gathered about individual pupils in the course of a school year has much more accuracy.

Classroom assessment places no political pressure on teachers; it can address the full range of learning goals. The low-pressure environment allows the teacher to ask potentially ambiguous questions that elicit higher-order thinking. In classroom tests, teachers can continue to reach for the conceptually more important questions because they have the luxury of responding to student questions during the assessment, allowing them to restructure the problem as they understand it, and accepting more than one answer as correct. In contrast, accountability tests cannot be experimental in nature; questions aimed at more ambitious constructs are left out if they might not withstand legal scrutiny.

The differences between accountability and instructional assessment are so fundamental and necessary that it may not be desirable to merge the two purposes. Instead, efforts to pursue an agenda like Glaser's, to develop formal measures of integrated understandings, and to train teachers in informal methods will have greater integrity if such procedures are *not* turned into accountability devices. While the substance of accountability tests must be improved by keeping conceptions of real learning in mind, their susceptibility to distortion should not be allowed to contaminate classroom uses of student data.

**Teaching to the test cheapens instruction and undermines the authenticity of scores as measures of what children really know.**

**Assessment tasks should be redesigned to more closely resemble real learning tasks.**

### Safeguarding Accountability Assessment

To restore the credibility of accountability tests, we must remove both the incentives and the means to distort scores. For example, rewarding high test scores with bonuses to schools or merit pay to teachers is a clear invitation to teach to the test. Because teaching to the test can raise scores more dramatically than can instruction designed to improve achievement (Shepard 1988), the incentive system could reward the worst practices.

We must institute procedural safeguards to signify that scores will not be used punitively and thereby protect the meaning of the test data. For example, the use of scientific sampling to report for an entire state reduces the threat to individual classrooms, because classroom results are not reported. Similarly, a fall testing date removes the ownership for scores from the teachers who supervise test administration, thereby reducing the incentive to redirect instruction to the test. When sampling is used, policymakers relinquish the ability to rank every teacher in the state by test scores but gain believable data. If school and district results are announced to the media, data should also be reported on the wealth of each community and the skills of incoming students, to avoid bragging by rich districts and punishment of poor districts.

The most important safeguard is not to give the same test repeatedly, year after year. Although using alternate forms of a minimum skills test does not prevent the teaching of too narrow a set of goals, new forms at least preclude rehearsing the specific vocabulary or word problems found on any one form of a test.

Another procedure designed to protect breadth of content is matrix sampling, used by the California Assessment Program. Every pupil at a grade level takes a test that is in fact a randomly assigned subpart of a larger test. Scores cannot be computed for individual pupils. However, the resulting scores reported for schools, districts, and the State of California are based on a broad content domain, making it less likely that teaching to the test will limit instruction to a narrow set of skills. Carrying this idea further, large-scale assessments could cover many broad domains by using a cycle of topics so that different subject areas could be comprehensively assessed each year.

**The dimensions of the test domain must be expanded so that teaching to the test does not imply teaching only a subset of learning goals.**

Ironically, the National Assessment of Educational Progress began almost 20 years ago with many of these substantive and procedural features, including the assessment of 10 subject areas in 4-year cycles. These features have gradually been altered in response to cost constraints and the shift from a monitoring purpose to accountability. The proportion of performance tasks compared to paper-and-pencil measures has decreased. The number of subject areas has been reduced in favor of more frequent assessments in the basics. Originally, data were reported at the item level, of use only to curriculum experts. Today results are aggregated into a single score for each subject area to produce a national report card. In 1968 a sampling design based on states was purposely avoided to prevent invidious comparisons and to ensure local cooperation with the assessment. In keeping with today's climate, however, state-by-state comparisons have been mandated by Congress beginning with the 8th grade mathematics assessment of 1990.

**The more we focus on raising test scores, the more instruction is distorted, and the less credible are the scores themselves.**

**Expanding Boundaries**

Accountability testing in the 1980s is having a pernicious effect on education. Standardized tests have always been fallible, limited measures of learning goals. Today, the joint effect of limited content and inordinate weight attached to test results is magnifying the drawbacks. Tests are being asked to do too much. The more we focus on raising test scores, the more instruction is distorted, and the less credible are the scores themselves.

The content boundaries of current assessments should be expanded to include conceptual understanding and problem-solving abilities. However, substantive improvements alone will not correct current ills. If policymakers want assessment data to reflect accurately what students know, they must remove incentives to teach to the test.☐

## *References*

Cannell, J.J. (1987). *Nationally Normed Elementary Achievement Testing in America's Public Schools: How All Fifty States Are Above the National Average.* Daniels, W.V.: Friends for Education.

Carpenter, T.P., E. Fennema, P.L. Peterson, C. Chiang, and M. Loef. (April 1988). "Using Knowledge of Children's Mathematical Thinking in Classroom Teaching: An Experimental Study." New Orleans: Paper presented at the annual meeting of the American Educational Research Association.

Cole, N.S. (1988). "A Realist's Appraisal of the Prospects for Unifying Instruction and Assessment." In *Assessment in the Service of Learning: Proceedings of the 1987ETSInvitationalConference.* Princeton, N.J.: Educational Testing Service.

Darling-Hammond, L., and A.E. Wise (1985). "Beyond Standardization: State Standards and School Improvement." *The Elementary School Journal* 85: 315-336.

Glaser, R. (1988). "Cognitive and Environmental Perspectives on Assessing Achievement." In *Assessment in the Service of Learning: Proceedings of the 1987ETSInvitationalConference.* Princeton, N.J.: Educational Testing Service.

Levin, H.M. (March 1987). "Accelerated Schools for Disadvantaged Students." *Educational Leadership* 44: 19-21.

McNeil, L.M. (1988). "Contradictions of Control, Part 3: Contradictions of Reform." *Phi Delta Kappan* 69: 478-485.

Popham, W.J. (1987). "The Merits of Measurement-Driven Instruction." *Phi Delta Kappan* 68: 679-682.

Rothman, R. (October 26, 1988). "Vermont Plans to Pioneer with 'Work Portfolios.'" ***Education Week*** 8, 1: 11.

Salmon-Cox, L. (September 1982). "MAP Math: End of Year One Report." Pittsburgh: Learning Research and Development Center.

Salmon-Cox, L. (September 1984). "MAP Reading End of Year Report." Pittsburgh: Learning Research and Development Center.

Shepard, L.A. (April 1988). "Should Instruction Be Measurement-Driven? A Debate." Paper presented to the annual meeting of the American Educational Research Association, New Orleans.

Tyson-Bernstein, H. (1988). "America's Textbook Fiasco: A Conspiracy of Good Intentions." *American Educator* 12: 20-27, 39.

# I-3

# Why Is Assessment Changing?

JUDITH A. ARTER
NORTHWEST REGIONAL EDUCATIONAL LABORATORY

LINDA BOND
NORTH CENTRAL REGIONAL EDUCATIONAL LABORATORY

The story of education in the '90s continues to be one of reform and change. In both the old educational system and much of the existing system, teachers provided knowledge and students received that knowledge. New thinking about curriculum, instruction, and learning requires us to pay attention to new ways of assessing student learning. Today, for example, we know that good instruction engages students in the learning process. What is important to learn is also changing. In the past, knowing a large number of facts was important; today, with information doubling every three years, students need to know how to access information and apply it to real-life situations. Students will face many situations that have no clear-cut, right answers; they will have to analyze situations and apply their knowledge and skills to find workable solutions.

Students still need knowledge and facts, but the emphasis is changing. Much of traditional assessment measured facts and skills in isolation, less frequently requiring students to apply what they know and can do to real-life situations. As the curriculum and instruction of the schools change, so too must the way we assess student learning.

Lack of change in assessment can get in the way of reform. For example, good teachers who change how they teach students and what they emphasize in the curriculum are often judged by assessment results that show how well their students have memorized facts and can recognize right answers. What kind of reinforcement do teachers receive for making the necessary changes in classroom practice if the outcomes assessed by the tests do not cover the full range of our desired outcomes?

For all these reasons, educators and educational researchers are working to change their assessment of student achievement to reflect changes in the expectations and standards for students, in the methodologies and purposes of assessment, and in the instruction that assessment is intended to support.

We call this evolving type of assessment alternative assessment. Alternative assessment includes any type of assessment in which students create a response to a question rather than choosing a response from a given list (e.g., multiple-choice, true/false, or matching). Alternative assessments can include short-answer questions, essays, performances, oral presentations, demonstrations, exhibitions, and portfolios.

---

Copyright © 1993 NWREL.

Even though we emphasize reasons for alternative assessment, we do not want to imply that only alternative assessments are worthwhile and all previous assessments are worthless. Actually, no single assessment approach can accomplish all our goals. We need to build a balance by considering all the things we want to assess and the reasons we want to assess them, and then choosing the assessment approach that best matches our targets and purposes. Sometimes the answer will be an alternative assessment, and sometimes the answer will be a multiple-choice test.*

## Current Changes and Their Effects

### Rationale for Changes in Assessment

Changes in assessment occur for a number of reasons, including changes in student standards, changes in valued outcomes, changes in the relevancy of assessment tasks, changes in the purposes of assessment, and changes in the roles of teachers in assessment. For all these reasons and more, educators and educational researchers are looking for ways to improve the alignment of assessment content and strategy with proposed changes in curriculum and instruction.

### Changes in Valued Outcomes

Goals for students are changing. For example, the National Council of Teachers of Mathematics (Commission on Standards for School Mathematics 1989) has released a set of national standards for mathematics, and several groups are drafting standards for science education (National Research Council 1994). All subject areas include an increasing emphasis on critical thinking, problem solving, metacognition, students' efficiency with tasks, group collaboration, reasoning, and lifelong learning. These new outcomes are difficult to measure in a multiple-choice format, and alternate ways to assess student competency are being sought.

### Changes in Content and Performance Standards

Definitions of competency and expertise are also more refined. For example, the definition of what it means to be a good reader has changed; just being able to convert symbols on the page to sound is not enough. To be functionally literate in tomorrow's world, students must draw inferences, relate current information to past information, have a variety of efficient reading strategies, and know when they do not understand and how to get help. These skills are difficult to measure in fixed-response format, because testing only skills is insufficient; we must also test a student's ability to know how and when to use a variety of skills collectively to accomplish a goal. For this reason, multiple sources of assessment information are needed, including assessments that allow for the evaluation of complex skills used together.

> . . . our understanding of reading and reading problems has changed in the past decade. What we used to consider 'sure signs' of reading difficulties—letter reversals, for example—may no longer be reliable, and other important signs may go unnoticed without a reconsideration of what we're looking for. Habit affects all of us. In classrooms and clinics, hundreds of thousands of teachers are approaching reading problems with methods based on old ideas that are now considered off the mark (Gillet and Temple 1990, p. 496).

### Changes in What We Want to Communicate

Teachers, parents, and students understand what is valued in education by virtue of what is assessed and how it is assessed. Using only multiple-choice tests gives the impression, for example, that only one right answer exists and that a right answer is always possible. Assessments also communicate who has the right and the ability to judge quality. Kober (1993) notes the following messages from traditional science tests:

- Because they have a single right or wrong answer, they reinforce the misleading conception of science as a static body of facts;
- Ranking individual performance against that of a larger group perpetuates the notion that only a few students—the top scorers—are smart enough to pursue science; and
- By sampling a breadth of content in a superficial and unconnected way, traditional tests actually reward instruction that drills students on low-level facts and name recognition.

Current theories seek to change the power relationships in the classroom, with teacher as facilitator rather than imparter of knowledge and students as assessors of their own and each others' performances. We need to reinforce these practices by making them part of assessment tasks and scoring criteria.

---

*Interestingly, 60 years ago multiple-choice tests were considered an alternative to the current assessment of the time—teacher-based judgment. Now "alternative" is used to mean assessments other than multiple-choice, true/false, and matching. Objectivity, however, still has the meaning it did 60 years ago. The goal now is to make subjective assessment (performance and other alternative assessment) as objective as possible.

### Changes in the Purposes of Assessment

In the past, assessment was frequently used for sorting: grading, selection into special programs, assignment to instructional groups, identification of the highest- and lowest- performing students, and so forth. Different skills were expected of students depending on their ability. Today, all students need to achieve high standards of performance to succeed in the world, and sorting can no longer be the primary goal of assessment.

> The new paradigm of educational evaluation is constructed around the new expectations of education in contemporary society. The old paradigm assumed that educational opportunities had to be rationed. . . . The new paradigm . . . recognizes that contemporary society is realizing that its members need increasing amounts of school learning in order that they may become . . . effective citizens,. . . helpful members of their communities and constructive participants in the economic life of society. . . . The role of educational evaluation in this connection is to help identify the assets of students on which effective educational programs can be built and the characteristics which are likely to interfere with the student's learning so that the school or college may help the student overcome these difficulties (Tyler 1986, p. 33).

Knowing that one student ranks higher than another is less important than knowing how both students' performances compare with our ultimate goals for performance. The assessment must describe student performance well enough so that students and teachers can assess progress toward agreed-upon goals. The standard is no longer "doing better than the rest."

### Changes in the Importance of Relevancy

Multiple-choice tests are stand-ins for actual performance. For example, we could use a multiple-choice test to assess knowledge about how to perform a science experiment. If the student does well on the test we infer that the student could do the actual experiment. Although many students who do well on multiple-choice tests of scientific knowledge also do well in the laboratory setting, the correlation does not hold for all students or for all occasions. To use a multiple-choice test as the exclusive measure of a student's understanding of scientific principles ignores the need to assess the application of that understanding in the laboratory or some other real-life setting that requires scientific reasoning.

Modern theories of teaching and learning have demonstrated that students know, understand, and retain more when they learn in the context of real-life situations. Students can also demonstrate depth of understanding when their task mirrors a real-life situation. Educators recognize the need to give students more realistic tasks to do on assessments, because we want to know if students can apply their knowledge to real-life situations and because we don't know if artificial situations allow students to show the full range of what they can do.

### Changes in the Role of Teachers in Assessment

In addition to changes in large-scale assessment, educators are looking anew at classroom assessment and classroom uses of large-scale assessment. Although assessment occurs frequently in the classroom, teachers rarely have the opportunity to learn the skills for assessing students well and how to use the assessment information to improve instruction. Teachers are the heart of educational improvement. They will administer and interpret the results of alternative assessments. If teachers don't feel this information is useful, or understand how it can be used to improve student learning, their resistance may cause the same contamination of assessment results that occur with more traditional large-scale assessments.

Alternative assessment is not magic. Teachers must understand the principles of assessment to ensure that assessments are not misused or underused. We must build in those features that make such assessments useful in the classroom, and then help teachers understand how to use them to improve their students' learning (see Section IV for papers on alternative assessments as instructional tools). If we don't, all the wonderful "authentic" information our alternative assessments deliver may not materialize.

The changes in education that matter occur in classrooms, in the day-to-day interactions between teachers and students. Properly designed alternative assessments can support changes in these interactions. Our knowledge of how students learn is increasing, and our assessments should reflect our new knowledge about teaching and learning.

### The Need for Professional Development

Most teachers have no training in assessment, even though researchers have shown that teachers use assessment information once every three minutes to make decisions about the direction of their instruction (Stiggins 1991). Teachers must understand how to make those observations systematically and with the necessary quality that ensures the assessment information is accurate, reliable, and matched to instructional goals. Teachers should not be expected to take on this increased role in assessment without support.

## Conclusion

Assessment is changing because education is changing. What students need to know and to be able to do is sometimes quite different from what schools now teach. Few aspects of society have been able to keep up with the rapid changes in our world. But keep up we must, if our students are to succeed in their changing world. Changes in assessment can promote and support necessary changes in curriculum and instruction.

### REFERENCES

Commission on Standards for School Mathematics. (1989). Curriculum and Evaluation Standards for School Mathematics, Reston, Va.: National Council of Teachers of Mathematics.

Gillet, J.W., and C. Temple. (1990). *Understanding Reading Problems*. 3rd ed. Glenview, Ill.: Scott Foresman/Little, Brown Higher Education.

Laboratory Network Program. (1993). *Improving Science and Mathematics Education/A Tool Kit For Professional Developers: Alternative Assessment (Pilot Test Version)*. Portland, Oreg.: Northwest Regional Educational Laboratory.

Kober, N. (1993). *Ed Talk: What We Know About Science Teaching and Learning*. Washington, D.C.: CEDaR.

National Research Council. (1994). "National Science Education Standards—Draft." Washington, D.C.: National Committee on Science Education Standards (phone: 202-334-1399).

Stiggins, R.J. (March 1991). "Assessment Literacy." *Phi Delta Kappan* 72, 7: 534-539.

Tyler, R. (1986). "Changing Concepts of Educational Evaluation." *International Journal of Educational Research* 10, 1: 1-113.

---

This chapter is adapted from Chapter 3 of *Improving Science and Mathematics Education/A Tool Kit for Professional Developers: Alternative Assessment* (Pilot Test Version), September 1993, 3.1–3.4. Developed by the Laboratory Network Program—Alternative Assessment. Available from Northwest Regional Educational Laboratory.

# 1-4

# Assessment

David T. Conley
University of Oregon

It may appear a bit jarring to separate the discussion of outcomes from that of assessment by the two intervening sections on curriculum and instruction. A case can certainly be made that there is a very close linkage between outcomes and assessments, which there is. At the same time, the two should not be confused. An assessment determines achievement of an outcome, and there can be many ways to do this. The outcomes themselves, however, serve to drive decisions about curriculum, instruction, and other related processes. The process of identifying outcomes, in this framework, precedes these other decisions, including the choice of assessments. In reality, this is rarely the case, since curriculum, instruction, and assessment already exist.

Once standards have been established and agreed upon, discussions of the proper assessment tools can take place in a more informed environment. Assessment can be altered so that it provides useful information to teacher, student, and parent about performance relative to district (or state) performance outcomes, to district curriculum objectives, and to individual learner goals. This rational, linear approach is rarely followed in such a step-by-step manner, in part because outcomes and assessment practices exist *de facto*, even if they are not written out and adopted formally. The fact they already exist means that any changes in these areas have immediate, direct implications for teaching and learning practices.

If the types of changes in curriculum and instruction that have been described in preceding sections take hold in American education, there will be a concomitant change in the philosophy and technology of assessment and in the outcomes identified as having primary importance for all children. The familiar saying "what gets measured gets done" has significance for school restructuring.

## Some Limits of Current Assessment Approaches

The current testing technology, from the classroom to the national level, is built largely on assumptions about goals of the curriculum and methods of instruction that are being questioned. Most current tests are designed to provide a summation of the factual information retained by students. These methods of testing are based on considerations of efficiency and have close ties to the behaviorist notion that all learning can be disaggregated into a series of measurable units and that the sum of the performance on the subunits accurately measures the full scope of what is known by the student. These tests may have high reliability (that is, perform similarly in different situations) but low validity (they may not measure what those giving the test want to know about the learner); in other words, they tell us

---

Reprinted with permission from D.T. Conley, Chapter 12 (pp. 151-164) of *Roadmap to Restructuring: Policies Practices, and the Emerging Visions of Schooling* (Eugene: ERIC Clearinghouse on Educational Management, University of Oregon, 1993).

a great deal about things that may have very little to do with what students actually know and are able to apply. They provide little insight into the ways in which discrete pieces of information are combined, or integrated, by the learner to solve real-world problems or to serve as the stepping stones to new learning experiences.

O'Neil (1992b) describes how behaviorist notions affected the ways in which instruction has been organized and assessed:

> Popularized by B.F. Skinner and others, the behaviorist view of learning, when translated by schools, was characterized by lengthy lists of measurable behavioral objectives and tightly sequenced curriculums. Knowledge and skills were broken down into smaller and smaller bits, under the assumption that mastering simpler steps would add up, in the end, to complex thinking. These "bits" tended not to be placed in the context of an authentic problem situation, and students had difficulty applying what they had learned in new contexts. Little attention was given, moreover, to the conceptions and misconceptions that learners held about the skills or knowledge being introduced; so misconceptions frequently resurfaced after the learning task concluded (p. 4).

These concerns are confirmed by the results from the three-year $1 million study sponsored by the National Science Foundation and conducted by researchers at Boston University (Rothman, October 21, 1992). The study reviewed standardized achievement tests and the tests contained in textbooks through the use of three strategies: (1) an item-by-item analysis of the six most widely used standardized achievement tests along with a sample of textbook tests in science and math in grades 4, 8, and high school; (2) a questionnaire administered nationwide to 2,229 math and science teachers in grades 4-12; and (3) interviews with 199 math and science teachers and 90 building-level administrators in six urban districts. The researchers found that those tests emphasized thinking and content at the levels of knowledge and comprehension. Very few questions measured conceptual knowledge, problem-solving, or other forms of higher-order thinking.

## Some Early Attempts to Develop New Assessment Tools

If expectations for student learning shift from measuring what students can repeat to demonstrating what they can do, the technology of testing can be expected to shift or evolve accordingly. This process of developing new methods is still in its infancy and will take a number of years to mature. In the meantime, schools involved in restructuring will be hard pressed to demonstrate improvement. The Saturn School in St. Paul, Minnesota, is an example of one such program caught in the middle (Weisman, July 31, 1991). The heralded model program found its students' scores on the Iowa Test of Basic Skills declined each of the first three times students took the test, including a ten-point drop in math scores over two years.

Since mathematics tests are the most fact-based and decontextualized of all achievement tests, this is the area where test scores would be expected to drop if a curriculum were decoupled from a traditional objectives-based structure and allowed to move more toward student-directed learning. Supporters of the Saturn School note that much of the computer technology necessary for the school's skills instruction program arrived late or was inappropriate for the program and had to be redesigned. Written accounts of the program do not indicate, however, that there was any sustained effort to develop alternative means of assessment to demonstrate what students did know. Schools such as this may well need to identify learner outcomes and develop assessment methods in tandem with creating new curriculum, time structures, or learning environments.

The difficulty of such an undertaking is highlighted by experts in the field of assessment, who suggest that five years may be a minimum period necessary to develop new assessment tools (Rothman, March 20, 1991). Eva Baker, codirector of the federally funded Center for Research on Evaluation, Standards, and Student Testing, believes it will take from five to ten years to develop and implement this new technology. She expresses concern over the ability of teachers to learn and implement new measurement techniques in the context of their existing responsibilities.

This strategy, however, assumes that development must be done by large research centers under the support of multimillion-dollar grants. While this sort of development effort may be valuable, particularly for creating alternatives to standardized achievement tests, it is likely that much, perhaps most, of the development work on new assessment techniques will take place at the state and district levels. The U.S. Department of Education is encouraging this effort through its support of the State Alternative Assessment Exchange, which is housed at the Center for Research on Evaluation, Standards, and Student Testing and cosponsored by the Council of Chief State School Officers. The exchange provides a central database and clearinghouse for alternative assessment strategies developed by states (Rothman, March 13, 1991).

For example, the California Assessment Program (CAP) has traditionally been based predominantly on standardized achievement tests. The new direction for the program "indicates the dedication of the California Department of Education to the development of assessments that challenge students to create performances and products that really matter" ("With Funds Restored, CAP Adds 'More Authentic Measures'" 1991, p. 2).

Plans for CAP in 1992 included the implementation of integrated language arts exams in reading and writing, and of new performance assessments in science. CAP has been employing writing assessments at the eighth-grade level since 1987. The program introduced open-ended math questions in 1989. Some examples of the skills and abilities students are expected to demonstrate through the new assessments include the following:

*Language Arts*

- Reflect the meaning-centered, literature-based curriculum described in the English-Language Arts Framework.
- Construct their own meanings, integrating new insights with the unique knowledge and experience each brings to the task.
- Integrate reading, writing, speaking, and listening in ways that are natural to good instruction.

*Mathematics*

- Respond to open-ended questions that:
  - Present students with a situation that is engaging.
  - Allow students at various levels of ability and experience to respond to problems with multiple entry points.
  - Encourage creative responses by permitting students to investigate several paths to a solution or find multiple solutions.
  - Direct students to write for an audience so they can demonstrate their abilities in effective communication.

*Science*

- Provide opportunities for students to find connections among scientific concepts and principles.
- Encourage students to discover and construct, through inquiry and investigation, the important ideas of science.
- Engage students in science thinking processes embedded in content.
- Enable students to move beyond the activity to apply knowledge and conceptual understanding.
- Allow students to demonstrate understanding by doing—by designing and performing investigations that ask them to observe, measure, classify, sort, infer, detect patterns, formulate hypotheses, and interpret results.

*History-Social Science*

- Enable students to demonstrate knowledge of history.
- Incorporate multicultural perspectives and interdisciplinary approaches, especially with art and literature.
- Encourage ethical understanding and civic virtue.
- Emphasize democratic values embodied in the United States Constitution and Bill of Rights.
- Promote knowledge and cultural awareness through study of history, the humanities, geography, and social sciences. (Adapted from "With Funds Restored..." 1991, p. 2)

The California state system, by virtue of its size, will have an impact on educational practices nationally. Publishers will not ignore changes in California, an important market for texts and tests. At the other end of the spectrum, Vermont is pioneering a statewide assessment program in writing and mathematics. Piloted initially during the 1990-91 school year in forty-eight schools, the assessment was scheduled to become the first statewide test in the state's history. All fourth- and eighth-graders were to be assessed in writing and mathematics in three ways:

- A uniform test, which uses equivalent tasks administered under the same conditions for each student. [The intent is to make these tasks performance-based.]
- A portfolio, which includes material collected during the course of the year.
- A "best piece," which represents what a student considers his or her best effort for the year. (Rothman 1990, p. 18)

Results from the portfolio and "best piece" assessments are to be reported in the form of a "school report day," in which parents and other community members will attend what amounts to a New England-style "town meeting" where the work of students will be reported, displayed, and discussed.

Rothman (1990) describes the content and structure of the writing and math portfolios:

> The writing portfolio is expected to contain poem, play, or personal narration; a "personal response" to a cultural or sports event, book, mathematics problem, or current issue; and prose pieces from classes other than English and language arts.

The material will be evaluated on at least seven criteria, including the degree to which the organization suits the writer's purposes, the writing exhibits a sense of personal expression, and the use of detail adds to clarity, as well as evidence of progress over time and evidence of opportunities for students to revise their work.

The math portfolio is expected to contain a range of materials that "demonstrate the student's ability to learn and understand materials beyond the 'facts and knowledge' level," according to a report by the math committee.

Such materials, the report states, could include: a solution to a problem assigned as homework; a problem made up by the student "with or without solution"; a paper done for another subject that includes math, such as an analysis of data presented in a graph; or entries from a journal. (p. 18)

School districts and schools are experimenting with methods of alternative assessment. This is not an unreasonable as it might first appear; if the results are not to be used for comparison between and among school buildings or districts, that is, if they are not *high-stakes* tests, then there is considerable latitude in the technical standards that need be applied to such development projects. If the data from the tests are going to be used by districts primarily for internal decision-making and for program improvement, then entirely different processes can be employed to develop these tests. Teachers can become much more centrally involved in assessment design to ensure that the results of assessments are of value to teachers.

This is not to say that standards of reliability and validity should be abandoned. In fact, teacher involvement can help improve validity in particular by helping to identify what needs to be assessed early in the design process. In the debate on assessment, it is becoming increasingly clear, however, that slavish adherence to the illusion that current assessment tools are objective and rational can be dangerous. Data from those assessments can be used as the basis for program decisions and, in some unfortunate cases, student academic placements. Test-makers are not necessarily to blame, since they often inform school personnel of the limitations of their instruments. And teachers, for their part, often find themselves in the unenviable position of either using marginally relevant test data or using no test data at all as the basis for making decisions. They cannot be faulted for selecting what often appears to be the lesser of two evils.

## Examples of Assessment Strategies

While large-scale projects are being undertaken by states, universities, and large research centers to develop new performance standards and new assessment tools and strategies, many educators at the district and school-site levels are actively involved in creating their own standards and assessment methods. They are not content to wait until the large-scale projects, with their long timelines for development and their tendency to produce a horse by committee, are ready for use.

Two school-based research projects (Rogers and Stevenson 1988) explored a variety of techniques for assessing student work. Assessment on a fifth-grade social studies unit included the following methods:

- *Small-group interviews*. In small group discussions with an adult, students are asked to explain what they have learned. The level of student understanding is probed and explored through these discussions.
- *Situational pictures*. Children view a picture of a situation that illustrates the conflict caused by the application of some right (a nativity scene on public property being taken down two weeks before Christmas), and are asked to discuss its significance and meaning.
- *Card sorts*. Students are provided information about key governmental roles and institutions and are asked to sort them into piles labeled "most important" and "least important," and to provide a rationale for the decisions they make.
- *Learning logs*. Students describe in a notebook the most important thing they learn each day, identify areas where they are confused, and so forth.
- *Leader snapshots*. Students view pictures of key government figures and then attempt to identify them and tell what they do.
- *Open-ended versions of conventional tests*. Students provide extended explanations to more traditional questions. After answering an agree/disagree question, students list examples and provide justification for agreeing or disagreeing. (pp. 69-70)

The alternative strategies used to assess student learning from an eighth-grade unit on the poet Robert Frost were of a very different nature. A series of longitudinal tests and interviews was employed. Beginning with a test given immediately following the completion of the unit, researchers returned periodically throughout the semester and readministered elements of the original test. They also interviewed students. The results of this procedure provided insight into what students actually retained over time from a unit where they scored well on the initial posttest. In addition, the assessment captured student perceptions and motivations related to the learning experience. This type of information is useful, not only to students, but to teachers, who can use the feedback the next time they prepare to teach the same material. Teachers often lack this type of information and mistakenly interpret the posttest results as an accurate gauge of student learning, as do the students.

A strategy that involves public demonstration of work by groups of students is exemplified by the Rural Educational Alliance for Collaborative Humanities (REACH) Program's use of an exposition for students from ten project sites to display their work (Barone 1991). The REACH Program encouraged students to explore their personal and community history and the culture of their rural community to help foster a sense of connection among the students, the school, and the community. Students produced writings, interviews, dramatic presentations, and media productions. These were presented at a two-day "exposition," along with portfolios demonstrating student work such as poetry, stories, and collections of essays that demonstrated the students' progress.

Portfolios of student work have been proposed as a potentially powerful tool for gauging student growth, encouraging self-analysis, and helping students to develop a sense of ownership and pride in their work. Paulson, Paulson, and Meyer (1991) provide guidelines for schools that are interested in using portfolios to reflect student progress:

1. Developing a portfolio offers the student an opportunity to learn about learning. Therefore, the end product must contain information that shows that a student has engaged in self-reflection.
2. The portfolio is something that is done *by* the student, not *to* the student. Portfolio assessment offers a concrete way for students to learn to value their own work and, by extension, to value themselves as learners. Therefore, the student must be involved in selecting the pieces to be included.
3. The portfolio is separate and different from the student's cumulative folder. Scores and other cumulative folder information that are held in central depositories should be included in a portfolio only if they take on new meaning within the context of the other exhibits found there.
4. The portfolio must convey explicitly or implicitly the student's activities; for example, the rationale (purpose for forming the portfolio), intents (its goals), contents (the actual displays), standards (what is good and not-so-good performance), and judgments (what the contents tell us).
5. The portfolio may serve a different purpose during the year from the purpose it serves at the end. Some material may be kept because it is instructional, for example, partially finished work on problem areas. At the end of the year, however, the portfolio may contain only material that the student is willing to make public.
6. A portfolio may have multiple purposes, but these must not conflict. A student's personal goals and interests are reflected in his or her selection of materials, but information included may also reflect the interests of teachers, parents, or the district. One purpose that is almost universal in student portfolios is showing progress on the goal represented in the instructional program.
7. The portfolio should contain information that illustrates growth. There are many ways to demonstrate growth. The most obvious is by including a series of examples of actual school performance that show how the student's skills have improved. Changes observed on interest inventories, records of outside activities such as reading, or on attitude measures are other ways to illustrate a student's growth.
8. Finally, many of the skills and techniques that are involved in producing effective portfolios do not happen by themselves....[S]tudents need models of portfolios, as well as examples of how others develop and reflect upon portfolios (pp. 61-63).

Another popular method of integrated performance assessment is holistic assessment. Generally applied to writing samples or student demonstrations in which it is important to consider the learning as a whole rather than as a series of component parts, holistic assessment generally relies on the use of a scoring rubric to determine student performance. The rubric contains specific descriptions of behaviors and evidence of performance an observer can use to analyze and categorize the student's performance along a continuum, usually designated by a numeric scale of 1 to 5, with five representing the highest, most competent, and most complex level of performance. One of the advantages of the rubric method of scoring is that it can be developed and applied by teachers. The behaviors identified as the focal point for observation in most rubrics are ones that can be grasped relatively easily by educators, students, parents, and community members. They also can generate discussion about what it is students should know, and at what levels and by what means they should demonstrate mastery of this knowledge.

An advantage of using rubrics is that they signal the outcomes necessary for success beforehand; learners don't have to guess what they must do to be successful. Furthermore, the rubric can be applied to preliminary drafts or be used throughout a course of study to provide formative feedback to the learner indicating clearly what he or she must do to improve performance. Such feedback can be more valuable and useful than a score of 64 to 72 on a test. The standard for success is identified before the fact, as well. A 3 on a scale of 1 to 5 might be designated as meeting the school's standards for mastery. Schoolwide profiles of student performance that are more descriptive than test scores can be developed and provided to teachers to help them pinpoint deficient areas to be addressed int he future. Schoolwide profiles also enable parents to understand what students can and cannot do as demonstrated by the assessment. This knowledge helps in the process of identifying school improvement goals.

Mark Twain Elementary School in Littleton, Colorado, created a rubric to judge written reports produced by fifth-graders as one element of an assessment process that also required them to research a topic, create a visual presentation relevant to the research topic, and deliver an oral presentation three to five minutes in length. Each element of the process was assessed separately, and a separate rubric was employed to assess the oral presentation as well. The written report was assesses employing the following five-point rubric:

> *5 - Excellent:* The student clearly describes the question studied and provides strong reasons for its importance. Conclusions are clearly stated in a thoughtful manner. A variety of facts, details, and examples are given to answer the question, and provide support for the answer. The writing is engaging, organized, fluid, and very readable. Sentence structure is varied, and grammar, mechanics, and spelling are consistently correct. Sources of information are noted and cited in an appropriate way.
>
> *4 - Very Good:* The student adequately describes the question studied and provides reasons for its importance. Conclusions are stated in a thoughtful manner, but with less clarity and insight than in an Excellent rating. A sufficient amount of information is given to answer the question, and provide support for the answer. The writing is engaging, organized, and readable. Sentence structure, grammar, mechanics, and spelling are generally correct, and sources of information are appropriately noted.
>
> *3 - Good:* The student briefly describes the question and has written conclusions. An answer is stated with a small amount of supporting information. The writing has a basic organization although it is not always clear and sometimes difficult to follow. Sentence structure and mechanics are generally correct with some weaknesses and errors. References are mentioned, but with some adequate detail.
>
> *2 - Limited:* The student states the question, but fails to fully describe it. The answers and/or conclusions given are vague, and basic information may be lacking. The writing generally lacks organization and is difficult to follow. There are many errors of sentence structure and mechanics. References may or may not be mentioned.
>
> *1 - Poor:* The student does not state the question. No answer or conclusion is given. The writing is disorganized and very difficult to read. Sentence structure and mechanics are consistently weak. References may or may not be mentioned.
>
> *0* - No written report is made. ("Mark Twain Elementary: The Peak Performance School," Littleton Public Schools)

The performance demonstration is yet another form of holistic assessment. Walden III, an alternative school in Racine, Wisconsin, with a long history of performance assessment, has developed what they title a "Right of Passage Experience" (ROPE). This process has served as a model for other schools. The model contains the following dimensions:

> All seniors must demonstrate mastery in fifteen areas of knowledge and competence by completing a **portfolio**, a **project**, and six other **presentations** before a ROPE committee consisting of staff members (including the student's home room teacher), a student from the grade below, and an adult from the community. Nine of the presentations are based on the materials in the portfolio and the project; the remaining six presentations are developed especially for the presentation process.
>
> *The Portfolio:* The portfolio, developed during the first semester of the senior year, is intended to be "a reflection and analysis of the graduating senior's own life and times." Its requirements are:
>
> 1. *A written autobiography*, descriptive, introspective, and analytical. School records and other indicators of participation may be included.
> 2. *A reflection on work*, including an analysis of the significance of the work experiences for the graduating senior's life. A resume can be included.
> 3. *Two letters of recommendation* (at minimum) from any sources chosen by the student.
> 4. *A reading record* including a bibliography, annotated if desired, and two mini-book reports. Reading test scores may be included.
> 5. *An essay on ethics* exhibiting contemplation of the subject and describing the student's own ethical code.
> 6. *An artistic product* or written report on art and an essay on artistic standards for judging quality in a chosen area of art.

7. *A written report analyzing mass media:* who or what controls mass media, toward what ends, and with what effects. Evidence of experience with mass media may be included.
8. *A written summary and evaluation of the student's course work in science/technology; a written description of a scientific experiment* illustrating the application of the scientific method; an *analytical essay* (with examples) on social consequences of science and technology; and *an essay on the nature and use of computers* in modern society.

*The Project:* Every graduating senior must write a library research-based paper that analyzes an event, set of events, or theme in American history. A national comparative approach can be used in the analysis. The student must be prepared to field questions about both the paper and an overview of American history during the presentations, which are given in the second semester of the senior year.

*The Presentations:* Each of the above eight components of the portfolio, plus the project, must be presented orally and in writing to the ROPE committee. Supporting documents or other forms of evidence may be used. Assessment of proficiency is based on the demonstration of knowledge and skills during the presentations in each of the following areas:

1. *Mathematics knowledge and skills* are demonstrated by a combination of course evaluations, test results, and work sheets presented before the committee, and by the ability competently to field mathematics questions asked during the demonstration.
2. *Knowledge of American government* should be demonstrated by discussion of the purpose of government; the individual's relation to the state; the ideals, functions, and problems of American political institutions; and selected contemporary issues and political events. Supporting materials can be used.
3. The *personal proficiency* demonstration requires the student to think about and organize a presentation about the requirements of adult living in our society in terms of personal fulfillment, social skills, and practical competencies; and to discuss his or her own strengths and weaknesses in everyday living skills (health, home economics, mechanics, etc.) and interpersonal relations.
4. *Knowledge of geography* should be demonstrated in a presentation that covers the basic principles and questions of the discipline; identification of basic landforms, places, and names; and the scientific and social significance of geographical information.
5. Evidence of the graduating senior's successful *completion of a physical challenge* must be presented to the ROPE committee.
6. A demonstration of *competency in English (written as well as spoken)* is provided in virtually all the portfolio and project requirements. These, and any additional evidence the graduating senior may wish to present to the committee, fulfill the requirements of the presentation in the English competency area.

The above is drawn from the 1984 student handbook, "Walden III's Rite of Passage Experience," by Thomas Feeney, a teacher at Walden III, an alternative public school in Racine, Wisconsin. Preliminary annotations are by Grant Wiggins (Cushman 1990, p. 10).

The preceding pages provide examples of both issues and techniques in alternative assessment. The process of developing such assessments is complex, yet critical to the success of school restructuring. Without clearly defined outcome standards and assessment methods, it will be difficult, if not impossible to demonstrate accountability in restructured learning environments. Parents, policy-makers, and the public at large appear less likely to tolerate a system of public education that lacks adequate accountability procedures and measures. In fact, public policy seems to be moving toward increasing accountability for public schools. This appears to be due at least in part to the continuing accretion from the local to the state level of the responsibility for financing local school districts as more courts find school financing schemes unconstitutional. Additional impetus for accountability comes from the expectation, held by a growing number of people, that governmental agencies and programs are responsible not just to provide services, but to effect outcomes, to make a difference, and to meet client needs if they are to justify their continued funding.

## The Challenge of Changing the Central Variables

Many restructuring schools establish some sort of project or change in school structure and identify this as restructuring. Others take a broader view and develop a vision, mission, and key strategic directions. Missing from both approaches is the development of performance standards and appropriate assessment methods—what its students will be able to do upon completion of an education and how such learning will be judged. As discussed earlier, performance standards can provide a framework within which appropriate assessment strategies can be selected or developed.

Once standards and assessment strategies are identified and agreed upon by a school faculty, the changes that need to be made in the structure of the school and the content and organization of the curriculum and instructional program may become much clearer. This process of working from outcomes to program structure makes it easier to determine which projects should be launched to restructure the school; it also establishes a framework within which all new ideas and proposals can be debated and analyzed by all faculty.

Such a conversation is not an easy one, particularly for faculty members who are not accustomed to discussing substantive issues related to curriculum, instruction, and assessment. Ultimately, such issues are rooted in values. Members of the school community must ask themselves: What is worth knowing? How is it best taught and assessed? If faculties and communities can reach some agreement on what young people should know, and how they should be expected to demonstrate their ability to apply what they know, the work of restructuring moves from a strategic level to a tactical one in which the primary question is: How can the school best be organized to ensure that the desired outcomes are achieved?

Such an approach challenges the traditional isolation of teachers. It suggests that teachers will have to teach in ways that will lead to certain agreed-upon outcomes. As schools currently exist, there are few mechanisms that help or expect teachers to think in terms of how their work relates to what other teachers are doing, have done, or will do. Performance standards arrived at through dialogue and discussion will tend to drive instruction in a school and will require considerable communication, cooperation, and collaboration among faculty to succeed. These norms may not be embraced with enthusiasm by teachers accustomed to working in isolation.

Furthermore, these discussions are difficult because they may call into question the value of a particular area of the curriculum or method of teaching. In some cases the threat goes even further, with certain existing subject areas not being included in lists of outcomes developed and approved. A process such as this is terribly difficult to conduct successfully in an environment such as a school (or school district) where there is a high premium on maintaining harmony, where many administrators pursue this goal over all others. The development of outcomes that result in accompanying changes in curriculum, instruction, and assessment will be very challenging for districts to do on a voluntary basis when motivated only by the goal of professional improvement of practice. Changes of this nature are highly political in nature and require a political context conducive to change. Elements of such a context will be considered in part 4.

The preceding four dimensions—outcomes, curriculum, instruction, and assessment—constitute the core of schooling. They are the areas that must be addressed sooner or later for substantial improvement or change in educational outcomes to be achieved. They are also the areas most difficult to change. This is true in part because many of the changes described in this chapter require teachers (and administrators) to examine many of their assumptions about learning and learners, about what constitutes valid educational experiences, and about how learning should be organized. Change is also hindered by the lack of resources for staff development and release time that would give teachers the opportunity to engage in the type of thoughtful analysis and planning necessary to rethink teaching and learning, and to develop new curriculum materials, teaching skills, and assessment strategies. Lack of adequate vision and leadership by those in administrative positions can make change of this nature much more difficult, or impossible.

# I-5

# Performance-Based Assessment in the Classroom: A Planning Framework

JAY MCTIGHE
MARYLAND ASSESSMENT CONSORTIUM

STEVEN FERRARA
MARYLAND STATE DEPARTMENT OF EDUCATION

Ongoing assessment of student learning is an essential aspect of effective teaching. Teachers can use a variety of assessment methods to diagnose students' strengths and needs, plan and adjust instruction, and provide feedback to students and parents regarding progress and achievement. We believe the primary purpose of classroom assessment is to inform teaching and improve learning, not to sort and select students or to justify a grade.

Although the choice of particular assessment methods should vary according to the purpose of the assessment, the content of the curriculum, and the age levels of students, we discuss a set of common principles underlying effective classroom assessment. We also consider the strengths and limitations of a variety of performance-based classroom assessment approaches and methods and present a series of vignettes to illustrate these methods in action. Finally, we offer a set of guiding questions and a framework for planning performance-based classroom assessments to improve teaching and learning.

## Principles of Effective Classroom Assessment

A wide variety of methods for assessing student learning is available to teachers (Airasian 1991, Cross and Angelo 1988, Ferrara and McTighe 1992). Regardless of the particular methods employed, effective classroom assessment is guided by three fundamental principles: (1) assessment should promote learning, (2) assessment should use multiple sources of information, and (3) assessment should provide fair, valid, and reliable information.

The first principle is based on the premise that the primary purpose of classroom assessment is to inform teaching and improve learning (Mitchell and

---

Reprinted with permission of the National Education Association from *Assessing Learning in the Classroom* (in press).

Neill 1992). This premise suggests that classroom assessment be viewed as an ongoing process instead of a single event at the conclusion of instruction. Rather than waiting until the end of a unit or course to assess students, effective teachers employ formative assessments at the beginning of instruction to determine students' prior knowledge. They assess regularly throughout the unit or course to obtain information to help them adjust their teaching based on the learning needs of students. They recognize that assessment results can inform them about the effectiveness of their teaching as well as the level of student learning.

When using performance-based assessments, teachers can make their evaluative criteria explicit in advance to serve as a focus for both instruction and evaluation. Effective teachers help their students understand that criteria describe the desired elements of quality (Ferrara, Goldberg, and McTighe 1995). They provide regular feedback to students based on the identified criteria, and they allow students to revise their work based on this feedback. They also actively involve students in improving their own performance by asking them to use the criteria to evaluate themselves and their peers.

Assessment for learning recognizes the mutually supportive relationship between instruction and assessment. Like a mobius strip, where one side appears to seamlessly blend into the other, classroom assessment should reflect and promote good instruction. For example, teachers following a process approach to teaching writing would allow their students to develop drafts, receive feedback, and make revisions as part of the assessment. Likewise, if we teach science through a hands-on, experimental approach, our assessment should include hands-on investigations.

A second principle of sound classroom assessment calls for a synthesis of information from several sources. A single assessment, such as a written test, is like a snapshot in that it provides a picture of student learning. Although a snapshot is informative, it provides incomplete information because it portrays an individual at one moment in time in one particular context. To use a one-time snapshot of student performance as the sole basis for drawing conclusions about how well a student has achieved desired learning outcomes is inappropriate. The classroom context offers a distinct advantage over large-scale assessments administered once or twice a year because it allows teachers to take frequent samplings of student learning using an array of methods. To continue the photographic analogy, classroom assessment enables us to construct a "photo album" containing a variety of pictures taken at different times with different lenses, backgrounds, and compositions. The photo album reveals a richer and more complete picture of each student than any single snapshot can provide. Applying the principle of multiple sources is especially important when the assessment information is used as a basis for important summative decisions, such as assigning report card grades or determining promotion.

A third principle of classroom assessment concerns validity, reliability, and fairness. Validity has to do with whether an assessment measures what it was intended to measure. For example, if a media specialist seeks to assess students' ability to conduct research using primary and secondary sources, she should observe their use of these sources directly as they work on their research projects. For this learning outcome, a paper-and-pencil test of students' knowledge of library references is an indirect and less valid assessment because it does not reveal the ability to actually use the references purposefully.

Reliability refers to the dependability and consistency of assessment results. If the same assessment yields markedly different results with the same students (without intervening variables such as extra instruction or practice time), one should question its reliability. Performance assessments present an additional challenge because they call for judgment-based evaluation of students' products and performances. A truly reliable evaluation results in equivalent ratings by the same rater on different occasions. For instance, an observation checklist can be used reliably as long as teachers are careful to ensure that their ratings will not differ substantially from occasion to occasion (e.g., Monday morning versus Friday afternoon). When teachers are involved in school- or district-level evaluations based on a set of criteria used throughout the school or district, inter-rater reliability must also be considered. In this case, scores on a writing assessment are considered reliable if different raters assign similar scores to the same essays.

Fairness in classroom assessment refers to giving all students an equal chance to show what they know and can do. Fairness is compromised when teachers assess something that hasn't been taught or use assessment methods that are incongruent with instruction (e.g., asking for recall of facts when the emphasis has been on reasoning and problem solving). The fairness of teacher judgments is also challenged by the "halo" and "pitchfork" effect, where expectations based on students' past attitudes, behaviors, or performances influence the evaluation of their current performance.

Subtle and unintended racial, ethnic, religious, gender, or other biases also present roadblocks to the fair assessment of students. Such biases may negatively influence students' attitudes toward, and performances on, classroom assessments. For example, the junior high mathematics teacher who routinely uses sports statistics as a main source for problem-solving tasks could irritate students who are not sports fans. Likewise, insensitivity to diverse religious beliefs (e.g., choosing reading passages involving only Christian holidays), gender and racial images (e.g., depicting all doctors as white males), or socioeconomic status (e.g., assuming that all kids have access to a telephone or home computer) may result in unfair evaluation of individuals or groups. Teachers must be on guard so that biases do not influence their evaluations of a student's performance.

Given these three general principles, we now consider a set of fundamental questions related to planning classroom assessments.

## Key Questions in Planning Classroom Assessments

Just as teachers can choose from numerous instructional techniques and strategies, so too can they select from a variety of methods for assessing learning. Figure 1 outlines some important questions that should influence the selection of assessment methods. The first question concerns learning outcomes, or the intended results of our teaching: *What do we want students to understand and be able to do?* Learning outcomes typically fall into three categories: (1) declarative knowledge—what we want students to understand (facts, concepts, principles, generalizations, (2) procedural knowledge—what we want students to be able to do (skills, processes, strategies), and (3) attitudes, values, or habits of mind—how we would like students to be disposed to act (e.g., appreciate the arts, treat people with respect, avoid impulsive behavior). The choice of specific assessment methods should be determined in large part by the nature of the learning outcomes being assessed (Marzano, Pickering, and McTighe 1993). For example, if we want students to demonstrate the capacity to write an effective persuasive essay, then our assessment should involve gathering samples of persuasive writing and evaluating them against specified criteria. In this case, a multiple-choice test would be ill-suited to measure the intended outcome. Likewise, if we want students to develop the ability to work cooperatively on a research project, then we should assess group processes and products as well as individual performance.

FIGURE 1
KEY QUESTIONS FOR CLASSROOM ASSESSMENT PLANNING

**Learning Outcomes**
*What do we want students to understand and be able to do?*

■ ____________________
____________________
____________________

■ ____________________
____________________
____________________

■ ____________________
____________________
____________________

■ ____________________
____________________
____________________

**Purpose(s) for Assessment**
*Why are we assessing and how will the assessment information be used?*

☐ to diagnose student strengths and needs
☐ to provide feedback on student learning
☐ to provide a basis for instructional placement
☐ to inform and guide instruction
☐ to communicate learning expectations
☐ to motivate; to focus student attention and effort
☐ to provide practice applying knowledge and skills
☐ to provide a basis for student evaluation (e.g., grading)
☐ to obtain data for site-based management
☐ to gauge program effectiveness

**Audience(s) for Assessment**
*For whom are the assessment results intended?*

☐ teacher/instructor
☐ students
☐ parents
☐ grade-level/department team
☐ other faculty
☐ school administrators
☐ curriculum supervisors
☐ business community
☐ higher education
☐ general public
☐ other: ____________________

Source: McTighe and Ferrara 1994

BEST COPY AVAILABLE

In addition to considering outcomes, we need to raise questions related to the purposes and audiences for classroom assessments: *Why are we assessing? How will the assessment information be used? For whom are the assessment results intended?* The purposes and audiences for assessments influence not only the methods selected, but the ways in which the results of the assessments are communicated. For example, if we wish to provide the parents of a primary-grade student with an interim report of progress in language arts, we might arrange a conference to give the parents a developmental profile of the child's reading skills and to review samples of the child's writing.

## Performance-Based Approaches to Assessing Student Learning

Given identified outcomes, purposes, and audiences, how might we assess student learning in our classrooms? The Framework of Assessment Approaches and Methods in Figure 2 offers a systematic guide to the purposeful selection of assessment methods. Each of the five columns in the framework identifies an assessment approach and contains examples of specific assessment methods corresponding to that approach. Given the focus of this chapter on performance-based assessment, we'll skip the first column (selected-response formats) and concentrate on the approaches in the second through fifth columns of the framework. We'll describe each general approach, examine the strengths and limitations of each, and provide vignettes of teachers using particular assessment methods in their classrooms.

Performance-based assessment refers to assessment activities that directly assess students' understanding and proficiency. These assessments allow students to construct a response, create a product, or perform a demonstration to show what they understand and can do. Because they call for students to apply knowledge and skills, rather than simply to recall and recognize, performance-based assessments are more likely to reveal student understanding than traditional assessments. They are well suited to assessing the application of content-specific knowledge, the integration of knowledge across subject areas, and lifelong learning competencies such as effective decision making, communication, and cooperation (Shepard 1989).

FIGURE 2

FRAMEWORK OF ASSESSMENT APPROACHES AND METHODS

**How might we assess student learning in the classroom?**

PERFORMANCE-BASED ASSESSMENTS

**SELECTED RESPONSE ITEMS**
- ☐ multiple-choice
- ☐ true-false
- ☐ matching

**CONSTRUCTED RESPONSES**
- ☐ fill in the blank
  - word(s)
  - phrase(s)
- ☐ short answer
  - sentence(s)
  - paragraphs
- ☐ label a diagram
- ☐ "show your work"
- ☐ visual representation
  - web
  - concept map
  - flow chart
  - graph/table
  - matrix
  - illustration

**PRODUCTS**
- ☐ essay
- ☐ research paper
- ☐ log/journal
- ☐ lab report
- ☐ story/play
- ☐ poem
- ☐ portfolio
- ☐ art exhibit
- ☐ science project
- ☐ model
- ☐ videotape/audiotape
- ☐ spreadsheet

**PERFORMANCES**
- ☐ oral presentation
- ☐ dance/movement
- ☐ science lab demonstration
- ☐ athletic competition
- ☐ dramatic reading
- ☐ debate
- ☐ musical recital

**PROCESS-FOCUSED**
- ☐ oral questioning
- ☐ observation ("kid watching")
- ☐ interview
- ☐ conference
- ☐ process description
- ☐ "think aloud"
- ☐ learning log

Source: McTighe and Ferrara 1994

BEST COPY AVAILABLE

## Constructed Response Assessments

Unlike selected-response items that call for a selection from given alternatives, constructed-response assessment tasks ask students to generate brief responses to open-ended questions, problems, or prompts. Short written answers and visual representations (e.g., concept map, flow chart, graph) are examples of widely used constructed-response assessment methods. Although constructed-response tasks may seek a correct or acceptable response (e.g., fill-in-the-blank), they are more likely to yield a range of responses. Thus, the evaluation of student responses requires judgment, guided by criteria. This approach may be used for assessing declarative knowledge and procedural proficiency. In addition, constructed responses can provide insight into understanding and reasoning when students are asked to show their work and explain or defend their answers in writing.

Constructed responses offer several advantages as classroom assessments. Because the response to each item is short, this assessment requires less time to administer than other types of performance-based assessments. A variety of constructed-response tasks can easily be used to assess multiple outcomes. Evaluation of student responses is fairly straightforward, guided by criteria or model responses.

Constructed responses are limited in their ability to adequately assess attitudes, values, or habits of mind. In addition, as with any open-ended assessment, judgment-based evaluation takes time and introduces potential problems of scoring reliability and fairness. For summative assessments, teachers should take care not to repeat constructed-response tasks so as to avoid having students give memorized responses to known questions and tasks.

### Classroom Examples

1. A middle school science teacher involves her students in an investigation of the absorbency rates of different brands of paper towels. Following the investigation, they record the results of their data collection on a chart and state their conclusions in writing. Students are evaluated on (a) the effectiveness of their charts in communicating results and (b) the accuracy of their written conclusions.

2. A 5th grade mathematics teacher asks her students to show their work as they attempt to solve multi-step word problems. In addition to examining their solutions, she looks at their work for evidence of the appropriate use of algorithms and problem-solving strategies. She provides feedback through brief written comments.

## Product Assessments

Student products provide tangible indicators of the application of knowledge and skills. Many educators believe that product assessment is especially "authentic" because the process closely resembles what students will encounter when they enter the work force. Teachers may evaluate written products (e.g., essays, research papers, laboratory reports), visual products (e.g., two- and three-dimensional models, displays, videotapes), aural products (e.g., an audiotape of an oral presentation), and other types of products to determine degrees of proficiency or levels of quality.

One application of product assessment occurs when representative samples of student work are systematically collected over time in portfolios. Portfolios allow teachers, students, parents, and others to observe a student's development and growth in learning. Over the years, portfolio assessment has been widely used in the visual arts, architecture, and certain technical areas, and now educators are witnessing a growing use of portfolios to document learning in other subject areas, especially the language arts.

The use of products and portfolios is appealing because of their instructional relevance. When students are given opportunities to produce authentic products, they often become more engaged in, and committed to, their learning. Unlike standardized assessments, which strive for uniformity, product assessments often present students with opportunities to express their individuality. Product assessments highlight what students can do, while revealing what they need to learn or improve. The criteria used to evaluate products, when made public, make the elements of quality known to students and serve as a guide for peer- and self-evaluation. Previously developed products can serve an instructional purpose when they are presented as models of excellence for students (Wiggins 1992).

Despite their benefits, product assessments have their drawbacks. Criteria for judging the products must be identified and agreed upon, and product evaluation can be a time-consuming process. When evaluating student products, teachers must be careful that their judgments aren't unduly influenced by relatively extraneous variables, such as neatness or spelling. Practicality must also be considered. The time required to develop high-quality products may compete with other instructional priorities. Product assessments require resources, including funds for materials and space for display and storage.

### Classroom Examples

1. Students develop a computer program for an advanced high school computer class. Their teacher evaluates students' programming knowledge and skills by examining the program's written code for accuracy and efficiency. In addition, students must run the program to demonstrate that it performs the

specified functions. Students must "debug" unsuccessful programs until they satisfy the requirements.

2. A 2nd grade teacher collects biweekly examples of representative student work in a language arts portfolio. During mid-year conferences, the teacher reviews the collected student samples with parents. The portfolio provides parents with tangible illustrations of their child's literacy development. The teacher uses the actual products, along with a developmental scale of reading and writing for the primary grades, to discuss the student's skill strengths and point out areas needing special attention.

3. Fifth grade art students create landscape paintings using tempera paints. Using a skills checklist, their art teacher assesses their paintings to determine their proficiency in using the medium. He also assesses their understanding of the use of compositional elements for creating an illusion of depth. Individual teacher-student conferences give students additional feedback on their work.

4. A middle school science teacher reviews her students' laboratory reports to determine how effectively students have applied the experimental procedures and collected accurate data. Her written comments in the margins point out errors and offer specific suggestions for improvement. The reports are returned, discussed, and then filed in the students' science folders for future reference.

### Performance Assessments

Using performance assessments, teachers are able to observe directly the application of desired skills and knowledge. Performance assessments can be among the most authentic types of student assessments because they can replicate actual performances required in the world outside of school. Performances have been widely used to assess learning in certain disciplines, such as vocal and instrumental music, physical education, speech, and theater, where performance is the natural focus of instruction. Teachers in other subjects, however, can also routinely include performances such as oral presentations, demonstrations, and debates as part of a broad array of assessment methods.

The evaluation of performances becomes instructionally valuable when students apply the scoring tools for peer- and self-evaluation. This kind of involvement helps students internalize the elements of quality embedded in the criteria. Many teachers have observed that students are motivated to put forth greater effort when they perform before "real" audiences of other students, staff, parents, or expert judges. In addition to the influence on students, schools often benefit from positive public relations when students perform for the community.

Despite their positive features, performance assessments can be time-consuming and labor-intensive for students and teachers. Time must be allocated for rehearsal as well as for the actual performances. The evaluation of performances is particularly susceptible to evaluator biases, making fair, valid, and reliable assessment a challenge.

#### Classroom Examples

1. Students in the school orchestra participate in a dress rehearsal two weeks before the public performance. The music teacher works with the students to evaluate their performance during the rehearsal and identify areas of weakness. During the ensuing practices, the orchestra members concentrate on making improvements in these areas prior to the actual performance before a live audience.

2. A high school social studies teacher sets up an in-class debate as a culminating activity for a contemporary issues unit. Students work as part of a team to debate the issue of gun control. The teacher will rate students' performances in the debates on several dimensions, including their understanding of the Bill of Rights, the persuasiveness of their arguments, their use of supporting factual information, their effectiveness in countering rebuttals, and their observance of the rules of debating.

3. An elementary phys ed teacher uses a skills checklist during the unit on introductory gymnastics to assess students' proficiency. Each student receives a copy of the checklist and works with a partner to try to perform the identified skills. The completed checklists constitute one component of the culminating grade for the unit.

4. A high school speech teacher works with a home economics teacher in preparing students to make oral presentations to communicate the results of a nutrition research project. Using a rating scale, the home economics teacher evaluates the students on the accuracy and completeness of their knowledge of the "food pyramid." The speech teacher uses a scoring rubric for the delivery of an informative speech to evaluate the oral presentations.

### Process-Focused Assessments

Process-focused assessments provide information on students' learning strategies and thinking processes. Rather than focusing on tangible responses, products, and performances, this approach seeks to gain insights into the underlying cognitive processes used by students. A variety of process-focused assessments are routinely used as a natural part of teaching. For example, teachers may elicit students' thinking processes using oral questions (e.g., "How are these two things alike and different?") or by asking students to think out loud as they solve a problem or make a decision. Teachers may also ask students to document their thinking over time by keeping a learning log. Finally, teachers can learn about students' thinking processes by observing students as they function in the classroom. This "kid watching" method is especially well suited to assessing the development of attitudes or habits of mind, such as persistence.

Process-focused assessments are formative in that they provide diagnostic information to teachers and feedback to students. They also help students develop metacognition by heightening their awareness of cognitive processes and worthwhile strategies. Process-focused assessment methods are typically used over time, rather than on single occasions. Thus, they are rarely used in standardized, high-stakes evaluations of students.

### Classroom Examples

1. A high school literature teacher regularly poses oral questions to assess students' interpretation of a text. Their responses sometimes reveal misunderstandings that need clarification by the teacher.

2. A kindergarten teacher interviews her students at the beginning of the school year. This informal assessment provides useful information about cognitive and linguistic development, social skills, and areas of personal interest.

3. A middle school social studies teacher carefully observes students to assess their cooperative skills as they work on a social studies project in learning groups. He also selects students to serve as process observers, giving them a checklist of observable indicators of cooperative skills. The teacher and student observers periodically provide feedback to the class on the effectiveness of their interactions in cooperative groups.

4. A high school mathematics teacher asks students to describe their reasoning processes by thinking out loud during the solution of open-ended problems. By listening to students as they articulate their thoughts, the teacher can identify fallacious reasoning, evaluate students' understanding of strategies, and then provide needed assistance.

## Evaluating Student Work and Communicating Assessment Results

In addition to making choices about classroom assessment methods, teachers should consider options for evaluating student work and for communicating assessment results. Figure 3 outlines some important questions to consider in deciding which options to use.

FIGURE 3

EVALUATION AND COMMUNICATION METHODS

**Evaluation Methods**

*How will we evaluate student knowledge and proficiency?*

**Selected Response Items:**

☐ answer key
☐ scoring template
☐ machine scoring

**Performance-Based Assessments:**

☐ generic rubric
☐ task-specific guide
☐ rating scale
- bipolar
- hierarchical

☐ checklist
☐ written/oral comments

**Evaluation Roles**

*Who will be involved in evaluating student responses, products, or performances?*

**Judgment-Based Evaluation by:**

☐ teacher(s)/instructor(s)
☐ peers
☐ expert judges (external raters)
☐ student (self-evaluation)
☐ parents/community members
☐ other ______________

**Communication/Feedback Methods**

*How will we communicate assessment results?*

☐ numerical score
- percentage scores
- point totals

☐ letter grade
☐ developmental/proficiency scale
- generic rubric
- task-specific guide
- rating scale

☐ narrative report (written)
☐ checklist
☐ written comments
☐ verbal report/conference

Source: McTighe and Ferrara 1994

BEST COPY AVAILABLE

## Evaluation Methods

*How will we evaluate student knowledge and proficiency?* The selection of evaluation methods should be determined largely by the assessment approach and the nature of the student responses to the assessment item or task. Performance-based assessments result in a wider range of responses, products, or performances that reflect different strategies, varying degrees of quality, and different levels of proficiency. We rely on judgment-based methods to evaluate responses to these open-ended assessments. Five primary types of evaluation methods are used in conjunction with performance-based assessments: scoring rubrics, task-specific scoring guides, rating scales, checklists, and written and oral comments.

### Rubrics

Because performance-based assessments generally do not yield a single correct answer or solution method but allow for a wide range of responses, evaluations of student products or performances are based upon judgments. The evaluative judgments are guided by criteria that define the desired elements of quality. One widely used scoring tool is a *rubric*, a generic scoring tool used to evaluate the quality of products and performances in a given outcome area. Rubrics consist of a fixed measurement scale (e.g., 4 points) and a list of criteria describing the characteristics for each point on the scale. Rubrics are frequently accompanied by examples of student products or performances to illustrate each of the points on the scale. These examples are called *anchors*.

Scoring rubrics can be *holistic* (intended to provide an overall impression of the elements of quality and levels of performance in a student's work), *analytic* (designed to indicate the level of performance of a student's work on two or more separate elements of quality), or *primary trait* (global in nature, like holistic rubrics, but focused on a specific feature, such as language usage).

### Task-Specific Scoring Guides

In contrast to generic rubrics, task-specific scoring guides are designed for use with a specific assessment activity (e.g., an open-ended question about a particular reading assignment or concept from a class discussion). Although they also contain a fixed scale and descriptive criteria, task-specific scoring guides cannot be used to evaluate responses to different performance tasks (Goldberg 1993).

Both rubrics and task-specific guides are most effectively used for evaluation or instruction when they are accompanied by examples of responses for each point on the scale. These examples, or anchors, provide tangible illustrations of the possible scores. Perhaps the greatest advantage of rubrics and task-specific guides lies in their capacity to clearly communicate elements of quality to students and evaluators. The clarity provided by criteria and anchors assists raters in reliably evaluating student responses, products, or performances. The criteria also provide targets toward which teachers can teach and students can aim. When students internalize the criteria contained in rubrics or guides, they are better equipped to evaluate and revise their own work.

These evaluation methods require time to collect or develop rubrics and task-specific guides, to identify representative anchors, to develop proficiency in applying them reliably, and to use them for evaluating student products and performances. Some schools and districts have recognized the significant professional development benefits of providing opportunities for teachers to work together on scoring student products and performances and identifying anchors.

### Rating Scales and Checklists

Rating scales and checklists are also helpful in evaluating responses to open-ended questions and tasks. Although these scoring tools are easy to use, they generally do not provide the detailed, explicit criteria found in rubrics and guides.

### Written and Oral Comments

Through written and oral comments, teachers can clearly and directly communicate with their students about elements of quality, expected standards of performance, areas of strength, and needed improvements. Although these methods are time-consuming, they allow teachers to provide specific evaluative feedback to students. The effectiveness of comments may diminish, however, if teachers provide only negative feedback (identifying errors or problems), make nonspecific positive comments that do not acknowledge particular aspects of student effort and work, or regularly make comments that do not address all the important elements of quality.

## Roles in Classroom Assessment

*Who will be involved in evaluating student responses, products, or performances?* As always, this guiding question should be answered with assessment outcomes, purposes, audiences, and methods in mind. The question also reminds us of the opportunity to involve others in the evaluation process. When students are engaged in applying criteria for self- and peer-evaluation, they begin to internalize elements of quality and performance standards in ways that can lead to improvements in the quality of their work and learning. Teachers may also involve other staff members, parents, or community experts in the evaluation of student products (e.g., science fair projects) and performances (e.g., public speaking exhibitions).

### Communication Methods

*How will we communicate assessment results?* A variety of methods can be used to communicate assessment results, including numerical scores, letter grades, developmental or proficiency scales, checklists, verbal and written comments or reports, and conferences. The choice of communication methods should be determined by the assessment purposes and methods, the evaluation methods, the feasibility considerations (e.g., time required), and, especially, the audience for the assessment.

## A Message to Students

Assessment is an essential component of the teaching-learning process. Without effective classroom assessment, it is impossible for us to know whether our students are hitting the target—that is, learning what is important for them to learn. The significance of classroom assessment, however, extends beyond the role of measuring learning. What we assess, how we assess and evaluate, and how we communicate results sends a clear message to students about what is worth learning, how it should be learned, what elements of quality are most important, and how well we expect students to perform. By considering the key questions and principles presented here, teachers will be better equipped to develop and use classroom assessments that provide fair, valid, and reliable information that will inform teaching and promote learning.

### REFERENCES

Airasian, P.W. (1991). *Classroom Assessment.* New York: McGraw-Hill.

Cross, K.P., and T.A. Angelo. (1988). *Classroom Assessment Techniques: A Handbook for Faculty.* (Technical Report No. 88-A-004.0). Ann Arbor: University of Michigan, National Center for Research to Improve Postsecondary Teaching and Learning.

Ferrara, S., G. Goldberg, and J. McTighe. (1995). "Ways in Which Teachers Communicate Performance Targets, Elements of Quality, and Standards of Performance." Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Ferrara, S., and J. McTighe. (1992). "Assessment: A Thoughtful Process." In *If Minds Matter: A Foreword to the Future,* Vol. 2, edited by A. Costa, J. Bellanca, and R. Fogarty. Palatine, Ill.: Skylight Publishing.

Goldberg, G. (1993). *Scoring MSPAP: A Teacher's Guide.* Baltimore: Maryland State Department of Education.

Marzano, R., D. Pickering, and J. McTighe. (1993). *Assessing Student Outcomes: Performance Assessment Using the Dimensions of Learning Model.* Alexandria, Va.: ASCD.

McTighe, J., and S. Ferrara. (1994). *Assessing Learning in the Classroom.* Washington, D.C.: National Education Association.

Mitchell, R., and M. Neill. (1992). *Criteria For Evaluation of Student Assessment Systems.* Washington, D.C.: National Forum on Assessment.

Shepard, L. (April 1989). "Why We Need Better Assessments." *Educational Leadership* 46, 7: 4–9.

Wiggins, G.P. (February 1992). "Standards, Not Standardization: Evoking Quality Student Work." *Educational Leadership* 4, 5: 18–25.

# I-6

# Understanding the Complexities of Setting Performance Standards

Robert J. Marzano
Mid-continent Regional Educational Laboratory

Arguing against the current trend to establish performance standards in content area classrooms is fairly difficult. After all, what could be more reasonable than to tell students what is expected of them? Why, then, do some teachers find setting performance standards so difficult? After working with hundreds of teachers in grades K-12 who represent virtually all content areas, I have concluded that difficulties arise in setting standards when we do not consider the basic nature of the knowledge being assessed. In fact, I believe that determining the type of knowledge that is being addressed is the first step to establishing clear performance standards.

## Two Basic Types of Knowledge

Most psychologists believe that knowledge comes in distinct types. Although many theories exist as to the nature of these types (e.g., Bloom et al. 1956, Gagne 1977, Merrill 1983), perhaps the most common classification scheme is declarative versus procedural knowledge. These two broad categories encompass all types of knowledge (Anderson 1983, Greeno 1978, Snow and Lohman 1989). Consider the categories of knowledge in columns A and B:

| **Column A** | **Column B** |
|---|---|
| • reading a map | • democracy |
| • performing long division | • a numerator |
| • editing an essay | • an amoeba |
| • shooting a free throw | • the conventions of punctuation |
| • setting up an experiment | • the rules of basketball |

Column A contains examples of knowledge that involve processes. Technically, this is referred to as procedural knowledge. Some procedures are performed in a linear fashion: For example, long division is a linear procedure. First you perform one step, then another, and so on. Other procedures require a more strategic approach: They might involve steps, but in no set order. In fact, the order used might vary from situation to situation. For example, reading a map can be described as a type of procedural knowledge that involves steps such as reading the name of the map, looking at the legend, considering the general scope of the area represented by the map, and so on. Yet these steps can be done in a variety of ways. Whether you decide to read the legend first or consider the general range of the area represented by the map would be considered your strategy, and it might change from situation to situation. When you acquire procedural knowledge, then, you learn specific steps or rules of performance and the manner in which they are best done.

The examples in column B do not involve processes or skills. Acquiring this type of knowledge involves understanding the component parts and to recalling or recognizing them. For example, acquiring the concept of democracy includes understanding that the people make decisions; each person has a single vote, votes are weighted equally, and so on. It also includes the ability to remember this information or at least recognize it later. This is called declarative knowledge.

An important distinction regarding these two types of knowledge is that performance standards look very different from type to type. A performance standard for procedural knowledge has different characteristics from a performance standard for declarative knowledge. Generally, performance standards are easier to construct for procedural knowledge than for declarative knowledge because of a close relationship between procedural knowledge and performance tasks.

## Performance Tasks and Procedural Knowledge

For the most part, performance tasks are a prerequisite to performance standards. A performance standard presumes some type of task that requires students to do something and produce something—to execute a process and to create a product. Marc Tucker, co-director of the New Standards Project, describes the relationship between performance standards and performance tasks in the following way: "You can't assess kids' performance unless you give them a task, and you can't assess their degree of achievement unless they actually perform the task" (1992, p. S3).

In addition to performance tasks' integral relationship with performance standards, many other characteristics make them attractive to educators. For example, Lauren Resnick asserts that performance tasks greatly expand the scope of what can be assessed. Whereas traditional tests provide assessment data regarding what information students know and what basic skills they can execute, performance tasks tell us how well students apply information:

> Many of the tests we use are unable to measure what should be the hallmark of a "thinking" curriculum: the cultivation of students' ability to apply skills and knowledge to real-world problems. Testing practices may, in fact, interfere with the kind of higher-order skills that are desired (Resnick 1987).

Other reasons commonly given for the use of performance tasks include:

- They provide clear guidelines for students about teacher expectations (Berk 1986).
- They reflect real-life challenges and situations (Hart 1994).
- They make effective use of teacher judgment (Archbald and Newmann 1988).
- They allow for student differences in style and interest (Mitchell 1992).

Although the use of performance tasks is one of the most popular topics in American education, the idea is not new. As Stiggins (1994) points out, performance tasks have been used for decades, at least, and possibly even centuries if one defines them in a broad sense. In tracing the recent history of performance tasks, Siegel (1986) explains that they were developed within domains involving technical "jobs" such as carpentry, welding, brick laying, lathe turning, and electronic assembly (p.127). To demonstrate competence in these areas, students performed the job-related skills. They used the lathe, laid bricks, welded, and so on. Obviously, much of what is done in these domains is procedural in nature. Laying bricks involves executing a set of skills and processes, as does using a lathe, and engaging in carpentry.

Performance tasks are quite compatible with content areas heavily laced with procedural knowledge. Consider mathematics. According to a recent study of the mathematics knowledge recommended by the National Council of Teachers of Mathematics, the National Assessment of Educational Progress, and the American Association for the Advancement of Science, many skills and processes are important to mathematics including: addition, subtraction, multiplication, division, constructing and converting ratios, decimals and percents, solving different types of problems, and the like (Kendall and Marzano 1994). Creating performance tasks within mathematics can be a simple matter of inserting the procedures inherent in that content area into realistic situations. For example, a performance task in mathematics might be structured around addition, subtraction, multiplication and division by asking students to determine how much material must be purchased and at what price to repave the playground or to build a new swing set within the a prescribed budget.

In summary, performance tasks for procedural knowledge require placing the target skills and processes in a realistic context. Once a performance task is designed, performance standards can be established.

## Establishing Performance Standards for Tasks Designed Around Procedural Knowledge

Establishing performance standards for tasks designed around procedural knowledge requires articulating expected levels of performance on the skills and processes embedded in the performance task. Commonly, a performance standard is embedded in a description of performance levels referred to as a rubric. Rubrics usually describe at least three levels of performance. The following is a generic three-level rubric for procedural knowledge:

> *Superior performance:* Carries out the major skills/processes inherent in the task with ease and automaticity.
>
> *Acceptable performance:* Carries out the major skills/processes inherent in the task without significant error, but not necessarily at an automatic level.
>
> *Inferior performance:* Carries out the major skills/processes inherent in the task with significant errors.

The performance standard in this rubric is the level titled "acceptable performance." If one receives a score or rating of inferior performance, then one has not met the performance standard; if one receives a score or rating of superior performance, then one has exceeded the performance standard. The key to making such a system useful in the classroom is to specify the exact nature of the skills and processes being assessed. To do this, one needs a thorough understanding of the nature and types of procedural knowledge.

### A Hierarchy of Procedural Knowledge

Procedural knowledge can be ordered hierarchically based on level of generality. At the top are highly robust and comprehensive processes with a diversity of possible products or outcomes and involve the execution of many interrelated subprocesses. Usually these subprocesses must be learned early, and to a level of automaticity or near automaticity. Then, the learner can spend the available capacity in working memory on the executive control of all the subprocesses.

Executive control involves regulatory activities such as deciding which subprocess is required and when, evaluating the functions of that subprocess, and planning the most appropriate activities to come next. High-level processes with these characteristics are macroprocesses (Kendall and Marzano 1994). For example, writing fulfills the defining characteristics of a macroprocess. Writing has a variety of possible products in that one can use the macroprocess to create expository discourse, narrative discourse, and so on. Also, to perform the macroprocess of writing, one must perform at an automatic or at least semiautomatic level such subprocesses as correcting faulty diction, correcting mechanical errors, controlling for logic of presentation, controlling for clarity of expression, and so on. Finally, even the most effective writer must expend a great deal of mental energy in the executive control of these subprocesses.

In the middle of the hierarchy are processes that cannot generate the variety of products possible from macroprocesses, and do not incorporate as wide a variety of subprocesses as macroprocesses (although they may involve some). These procedures are tactics or strategies. For example, an individual may have a tactic or strategy for solving a particular type of problem. Commonly, tactics and strategies are comprised of general rules. For example, most people have a tactic or strategy for reading a bar graph, which might include rules such as: (1) determine the elements depicted in the legend, (2) determine what is reported by one axis of the graph, (3) determine what is reported by the other axis of the graph, and (4) determine the relationship between the elements in the two axes. Although the sequence in which these steps occur contains a general pattern, no rigid set of steps exists. Consequently, some significant variation occurs in how different learners approach the process of reading a graph.

At the bottom of the procedural hierarchy are algorithms. These procedures normally do not vary in application, have very specific outcomes, and must be learned to the level of automaticity to be useful. For example, many computing processes in mathematics and decoding processes within reading are algorithmic in nature. Again, algorithms are frequently the subcomponents of macroprocesses and tactics.

### Emphasizing General Processes

Understanding this hierarchy of procedural knowledge is key to setting performance standards, because general types of procedures are more amenable to performance standards than specific types of procedures. Specifically, setting performance standards is a fairly straightforward process if macroprocesses are used. For example, if writing is the focus of a performance task, then the performance standard would specify subprocesses to be executed without significant errors. To illustrate, the performance standard might state that students perform the writing process without significant errors in the following writing subprocesses: controlling for overall logic; ensuring that sentences relate to each other; controlling for effective word choice; controlling for proper use of grammar. Targeting subprocesses for mastery at a particular age or grade level is key to setting performance standards. Whereas educators might expect high school students to have mastered the composing subprocess of controlling for overall logic, they would not expect the same of students in the primary grades.

In summary, when a macroprocess is the basis for constructing performance tasks, setting performance standards amounts to specifying expectations about the subprocesses within the macroprocess. Unfortunately, when a strategy or algorithm is the basis of a performance task, setting expectations about subcomponents can be a problem because many strategies and algorithms have no clear superordinate/subordinate structure of subcomponents. Rather, the subcomponents are individual steps that must be performed at the same level of accuracy and efficiency for the procedure to work at all. For example, which component parts of the process of three column addition would one expect students to execute with no errors and which component parts would one not expect students to execute? Could students add individual columns correctly, but make errors in carrying and still be judged as meeting the performance standard for three-column addition? Obviously not. Algorithms and strategies are comprised of fairly uniform sets of steps or heuristics, all of which must be executed at the same level of accuracy for the strategy or algorithm to be even moderately useful. For this reason, performance tasks are more effectively designed around macroprocesses than around strategies or algorithms.

## Creating Performance Standards for Tasks Designed Around Declarative Knowledge

Setting performance standards for tasks designed around declarative knowledge is more difficult than setting standards for tasks designed around procedural knowledge because of the nature of declarative knowledge and its lack of inherent relationship with performance tasks. Again, consider the hierarchical structure of declarative knowledge.

### A Hierarchy of Declarative Knowledge

As with procedural knowledge, declarative knowledge can be ordered hierarchically based on levels of generality. At the bottom of the hierarchy are facts about specific persons, places, things, and events; at the top are concepts and generalizations. For example, "John Kennedy was assassinated on November 22, 1963," is a fact. "People holding high political office put their lives in jeopardy" is a generalization. "Political assassinations" is a concept.

Although facts are important, generalizations and concepts help students develop a broad knowledge base because they transfer to different situations. Both the generalization about people in high political office and the concept of political assassinations can be applied across countries, situations, and ages, whereas Kennedy's assassination is a specific event not directly transferable to other situations. This is not to say that facts are unimportant. On the contrary, to truly understand generalizations and concepts, we must support them with exemplifying facts. For instance, to understand the generalization about people in high political office we need to understand some facts, one of which is probably Kennedy's assassination.

Facts versus concepts and generalizations represent two ends of the spectrum of declarative knowledge. In between are structures thought of as organized sets of facts; some are time sequences, causal networks, events, and episodes. (For a detailed discussion of the different types of declarative knowledge see Kendall and Marzano 1994.)

Just as performance tasks designed around procedural knowledge are most effective when organized around high-level macroprocesses, performance tasks designed around declarative knowledge are most effective when constructed around high-level declarative structures—in this case, concepts and generalizations. A performance task structured around a fact like "John Kennedy was assassinated on November 22, 1963" is ineffective. However, a performance task constructed around the generalization that "people holding high political office put their lives in jeopardy" or the concept "political assassinations" makes sense. The performance standard would then specify which facts, time sequence, causal networks, and so on would acceptably illustrate the concept of generalization. Selecting a high-level piece of declarative knowledge, then, has the same benefits as selecting a high-level piece of procedural knowledge in terms of constructing performance standards.

However, whenever one constructs a performance task for declarative knowledge—declarative knowledge does not inherently involve performances and products as does procedural knowledge, and, as we have seen, performances and products are central to performance tasks. No matter what type of procedural knowledge one identifies, it will automatically involve a performance and a product. For example, the use of the algorithmic procedure of multiplication always involves the performance or execution of the steps involved in the algorithm and it always involves the production of new information for the learner. The macroprocess of writing will always involve the execution of the embedded subprocesses and will always involve a product.

However, declarative knowledge is not inherently performance and product oriented. Understanding the concept of "political assassination" is shown by recalling or recognizing specific pieces of information and their relationship to each other. Unless one considers recalling and recognizing performances, declarative knowledge does not inherently involve a performance. Additionally, no new product is created when the concept of political assassination is recalled or recognized, because the information is in long-term memory of the individual being assessed or it could not have been recalled or recognized. No performance is executed and nothing new is created when one recognizes or recalls declarative knowledge.

Constructing performance tasks around declarative knowledge, then, requires the addition of a new component—mental operations that apply declarative knowledge and involve performances and products.

### Operations That Apply Declarative Knowledge

Because declarative knowledge does not inherently involve performances and products, one must identify operations that can be performed on declarative knowledge that inherently involves performances and products—mental operations that require one to apply declarative knowledge. Fortunately, researchers and theorists have generated a number of descriptions of such processes. Here we consider three models of knowledge application processes.

We can safely say that most teachers are familiar with Bloom's taxonomy of the cognitive domain. Recall that Bloom and his colleagues (1956) identified six levels of cognitive operations: (1) knowledge, (2) comprehension, (3) application, (4) analysis, (5) synthesis, and (6) evaluation. For the most part, Bloom's first two levels simply represent the acquisition of declarative knowledge. Levels 3 through 6 are where Bloom directly or indirectly addresses the issue of applying declarative knowledge.

Despite the rather straightforward nature of Bloom's taxonomy, research has shown that teachers have problems differentiating between the various levels of the taxonomy (Ennis 1981, Wood 1977). When asked to determine if a specific task uses the cognitive process of analysis or evaluation, teachers disagree more often than not. Despite confusion over Bloom's higher levels, the taxonomy has certainly been the educator's most widely used resource to identify ways to apply declarative knowledge.

Quellmalz (1991) has identified another useful categorization of cognitive processes. Stiggins (1994) describes these processes as:

*recall:* mastering foundational knowledge
*analysis:* determining ingredients or component parts
*comparison:* reasoning in terms of similarities and/or differences
*inferences:* reasoning inductively and deductively
*evaluation:* expressing and defending an opinion or point of view

Stiggins explains that everything above the recall level deals with the application of knowledge. Although little research exists on the utility of the Quellmalz framework, what is available is promising in terms of its utility as a framework for constructing performance tasks (Quellmalz 1991, Quellmalz and Hoskyn 1988).

A third framework commonly used to identify knowledge application processes was developed by Marzano and others (Marzano 1992; Marzano, Pickering, Arredondo, Blackburn, Brandt, and Moffett 1992; Marzano, Pickering, and McTighe 1993). This model describes 13 general reasoning processes that can be applied to any type of declarative knowledge. A useful characteristic of these reasoning processes is that they are easily cued by specific questions. (See Figure 1.)

Using Figure 1, a teacher would ask questions about the declarative knowledge the performance task was to be built around. To illustrate, consider the following generalization: "Historical events commonly happen in patterned ways." Using Figure 1, the teacher would run through the questions to see if any had potential as a useful way of applying the targeted declarative knowledge. The teacher might ask herself "Is there anything about which students should describe similarities and differences to demonstrate their knowledge of the generalization?" "Is there anything students should classify into groups to demonstrate their knowledge of the generalization?" and so on.

Based on these stimulus questions, the teacher might present students with a number of historical events that they had studied and ask them to classify those events into two or more categories, each of which exemplifies the generalization that "Historical events happen in patterned ways." Students might be asked to report on the categories they identified, the rules for category membership, and the extent to which their categories exemplify the target generalizations. This task would now have (1) a clear performance component—the execution of the mental process of classification, (2) a clear product—the students' description of their categories and the rationale behind them, and (3) a specific declarative knowledge component—the students' understanding of the generalization that "Some historical events happen in patterned ways."

In summary, to create a performance task around declarative knowledge, a teacher must select a mental process to be applied to the declarative knowledge and then ask students to report and describe the results of their applications. Three traits are necessarily involved in such a task: the product, the performance, and the declarative knowledge.

FIGURE 1.
QUESTIONS RELATED TO KNOWLEDGE APPLICATION COGNITIVE PROCESSES

| **Questions** | **Reasoning Process** |
|---|---|
| How are these alike? | Comparing |
| What groups can I put things into? What are the rules governing membership in these groups? | Classifying |
| What conclusions/generalizations can you draw from this, and what support do you have for these conclusions? What is the probability of this happening, and what support do you have for this conclusion? | Induction |
| What has to be true given the validity of this principle? What is the proof that this must be true? | Deduction |
| What's wrong with this? What specific errors have been made? How can it be fixed? | Error analysis |
| What is the support for this argument? What are the limitations of this argument? | Constructing support |
| What's the general pattern of information here? Where else does this apply? How can the information be represented in another way (graphically, symbolically)? | Abstracting |
| What do you think about this issue? On what do you base your opinion? What is another way of looking at the issue? | Analyzing perspectives |
| What/whom would be the best or worst? Which one has the most or least? | Decision making |
| What are the defining characteristics? Why/how did this happen? What would have happened if.. | Definitional, historical, and .projective investigation |
| How can I overcome this obstacle? Given these conditions, what should I do to accomplish the goal? | Problem solving |
| What do I observe? How can I explain it? What can I predict from it? | Experimental inquiry |
| How can this be improved? What new thing is needed here? | Invention |

(Adapted with permission from *Assessing Student Outcomes: Performance Assessment Using the Dimensions of Learning Model.* Marzano, Pickering, and McTighe 1993, ASCD. Copyright McREL Institute.)

## Setting Performance Standards for Declarative Tasks

Setting performance standards for declarative tasks forces a teacher to decide if the multiple traits involved will be assessed separately or in combination. In this section, we consider how performance standards are set when multiple traits are considered separately. In the next section, we will consider how performance standards are set when multiple traits are considered together.

In general, tasks designed around declarative knowledge will always have three performance standards—one performance standard for the declarative knowledge embedded in the task, one performance standard for the cognitive operation that applies the embedded declarative knowledge, and one performance standard for the product of the knowledge application. If one wished to design a performance standard for each of these traits, then three separate rubrics would be created—one rubric for each trait.

A generalized rubric for declarative knowledge embedded in a performance task might be stated in the following way. Using our sample task, this rubric would be applied to the generalization about historical events occurring in patterned ways.

> *Superior performance:* Displays an in-depth knowledge of the concepts, generalizations and facts embedded in the performance task. Demonstrates exceptionally broad and detailed knowledge of this information or demonstrates insights into this information that are not explicit in the information as presented.
>
> *Acceptable performance:* Displays a complete knowledge of the concepts, generalizations and facts embedded in the performance task. Has no significant misconceptions and is not missing any key information in his or her knowledge base.
>
> *Inferior performance:* Displays an incomplete or inaccurate knowledge of the concepts, generalizations and facts embedded in the performance

task. Has significant misconceptions or is missing some key features of the information.

In addition to a rubric describing various levels of understanding regarding declarative knowledge, one would also address the cognitive operation used to apply the declarative knowledge. Here is a generic three-point rubric that might be applied to any knowledge application cognitive operation:

> *Superior performance:* Performs the steps of the knowledge application process precisely, accurately, and with relative ease.
>
> *Acceptable performance:* Performs the steps of the knowledge application process without any significant errors, but not necessarily with relative ease.
>
> *Inferior performance:* Makes significant errors in the execution of the knowledge application process.

This rubric can be adapted to any cognitive operation used in a performance task. For the task involving the generalization about historical events occurring in patterned ways, the generic rubric would be rewritten specific to the knowledge application process of classification.

In an effort to be much more specific about the rubrics used for knowledge application cognitive operations, Marzano, Pickering, and McTighe (1993) have developed rubrics for the 13 processes described in Figure 1. In their model, the steps inherent in each process are identified and then rubrics designed for these steps. For example, the knowledge application process of classification is defined as involving the following steps:

1. selection of items or events to be classified
2. creation of useful categories
3. accurate and complete description of the rules for category membership
4. accurate placement of items or events into categories.

Marzano, Pickering, and McTighe present rubrics for each component of each process so performance on selected components can be assessed.

The third trait inherent in performance tasks designed around declarative knowledge is the product the description or demonstration of students' conclusions and findings. This trait can be assessed for the clarity of the presentation or demonstration. Below is a generic three-point rubric for clarity of presentation or demonstration:

> *Superior performance:* Clearly and effectively communicates the main idea or theme and provides support that contains rich, vivid, and powerful detail.
>
> *Acceptable performance:* Clearly communicates the main idea or theme and provides suitable support and detail.
>
> *Inferior performance:* Communicates information, but not a clear theme or overall structure.

In summary, whenever a performance task is designed around declarative knowledge, three traits can be assessed each requiring a performance standard and a rubric. Conversely, tasks designed around procedural knowledge can have only one performance standard and only one rubric. This innate difference between performance tasks based on procedural knowledge and those based on declarative knowledge highlights a major issue in the performance task literature and practice—the issue of holistic versus analytic scoring.

## Holistic Versus Analytic Scoring

Holistic scoring combines performance standards for multiple traits into a single rubric. Analytic scoring assesses traits separately. The previous example of the three rubrics used in the task about the generalization regarding historical patterns illustrates analytic scoring. With holistic scoring, a single rubric is used and its levels are composites of the various rubric levels for the individual traits embedded in the task. For example, a holistic performance standard for the task about the history generalization would include the performance standard for each of the individual traits. Unfortunately, this practice can create logical problems: Assume that a holistic task involved four traits: trait a, trait b, trait c, and trait d. A schematic of a three-point holistic rubric for such a task would be:

> *Level 3:* Level 3 performance for trait a; level 3 performance for trait b; level 3 performance for trait c; level 3 performance for trait d.
>
> *Level 2:* Level 2 performance for trait a; level 2 performance for trait b; level 2 performance for trait c; level 2 performance for trait d.
>
> *Level 1:* Level 1 performance for trait a; level 1 performance for trait b; level 1 performance for trait c; level 1 performance for trait d.

The logical problem with a holistic rubric occurs when a student has a highly uneven profile on the traits assessed by the task. For example, assume that a student exhibited the following profile on the four traits, a, b, c, and d:

*Trait a:* level 3
*Trait b:* level 3
*Trait c:* level 3
*Trait d:* level 1

Using the holistic rubric above in a strictly logical fashion, the student would receive a score of 1 even though he had exhibited level 3 performance on three out of the four traits. The logic of a holistic rubric demands that all traits at a given performance level are executed at least at that level. Stated in negative terms, a student's score on a holistic rubric must always devolve to the rubric level that represents the student's lowest level of performance in the multiple traits assessed by the rubric.

Logical problems with holistic rubrics has led some researchers, like Stiggins, to articulate strong warnings against their use:

> As a personal aside, I must say that I am minimizing my own use of holistic scoring. . . . I see few applications for such a score in the classroom. Besides, I have begun to question the meaning of such scores. I have participated in some writing assessments in which students whose analytic profiles of performance . . . were remarkably different ended up with the same holistic score. That gives me pause to wonder about the real meaning and interpretability of holistic scores. I have begun to think holistic scores mask the kind of more detailed information needed to promote classroom-level student growth (Stiggins 1994, p. 196).

I have concluded that analytic scoring should be used for any task that involves multiple traits, particularly when performance tasks are used for classroom assessment.

## Conclusions

Setting performance standards is a popular educational trend, but far more difficult and complex than its intuitive appeal to educators would suggest. Setting performance standards is intimately tied to performance tasks and the type of knowledge a performance task is designed to measure. If tasks are designed around procedural knowledge, the performance task and the performance standard are quite straightforward. However, if performance tasks are designed around declarative knowledge, at least three traits must be considered: the declarative knowledge embedded in the task, the application process applied to the knowledge, and the product or demonstration of the conclusion of the knowledge application. For such tasks, a separate rubric should be used for each trait embedded in the task. This takes time and a disposition to rigor. However, without taking adequate time and designing standards with sufficient rigor, one can quickly diminish the effectiveness and utility of performance standards.

*Note:* The author acknowledges John Kendall for the contribution his thinking and conceptualizing has made to this paper.

## REFERENCES

Anderson, J.R. (1983). *The Architecture of Cognition.* Cambridge, Mass.: Harvard University Press.

Archbald, D.A., and F.M. Newmann. (1988). *Beyond Standardized Tests.* Reston, Va.: National Association of Secondary School Principals.

Berk, R.A., ed. (1986). *Performance Assessment: Methods and Applications.* Baltimore, Md.: The Johns Hopkins University Press.

Bloom, B.S., M.D. Engelhart, E.J. Furst, W.H. Hill, and D.R. Krathwohl, eds. (1956). *Taxonomy of Educational Objectivities: The Classification of Educational Goals. Handbook I: Cognitive Domain.* New York: David McKay.

Ennis, R.H. (1985). "Goals for a Critical Thinking Curriculum." In *Developing Minds: A Resource Book for Teaching Thinking* (pp. 54-57), edited by A.L. Costa. Alexandria, Va.: ASCD

Gagne, R.M. (1977). *Conditions of Learning,* 3rd ed. New York: Holt, Rinehart and Winston.

Greeno, J.G. (1978). "A Study of Problem Solving." In *Advances in Instructional Psychology, Vol. 1,* pp. 13-75, edited by R. Glaser. Hillsdale, N.J.: Erlbaum.

Hart, D. (1994). *Authentic Assessment: A Handbook for Educators.* Menlo Park, Calif.: Addison Wesley.

Kendall, J.S., and R.J. Marzano. (1994). *The Systematic Identification and Articulation of Content Standards and Benchmarks: Update.* Aurora, Colo.: Mid-continent Regional Educational Laboratory.

Marzano, R.J. (1992). *A Different Kind of Classroom: Teaching with Dimensions of Learning.* Alexandria, Va.: ASCD

Marzano, R.J., D.E. Pickering, D.E. Arredondo, G.J. Blackburn, R.S. Brandt, and C.A. Moffett. (1992). *Dimensions of Learning Teacher's Manual.* Alexandria, Va.: ASCD

Marzano, R.J., D.E. Pickering, and J. McTighe. (1993). *Assessing Student Outcomes: Performance Assessment Using the Dimensions of Learning Model.* Alexandria, Va.: ASCD

Merrill, M.D. (1983). "Component Display Theory." In *Instructional-Design Theories and Models: An Overview of Their Current Status,* pp. 279-333, edited by C.M. Reigeluth. Hillsdale, N.J.: Erlbaum.

Mitchell, R. (1992). *Testing for Learning: How New Approaches to Evaluation Can Improve American Schools.* New York: The Free Press.

Quellmalz, E. (1991). "Developing Criteria for Performance Assessments: The Missing Link." *Applied Measurement in Education* 4, 4: 319-332.

Quellmalz, E., and J. Hoskyn. (1988). "Making a Difference in Arkansas: The Multicultural Reading and Thinking Project." *Educational Leadership* 45, 7: 51-55.

Resnick, L.B. (1987). *Education and Learning to Think.* Washington, D.C.: National Academy Press.

Siegel, A.I. (1986). "Performance Tests." In *Performance Assessment: Methods and Applications,* pp. 121-142, edited by R.A. Berk. Baltimore, Md.: The Johns Hopkins University Press.

Snow, R.E., and D.F. Lohman. (1989). "Implications of Cognitive Psychology for Educational Measurement." In *Educational Measurement,* 3rd ed., edited by R. Linn. London: Collier Macmillan Publishers.

Stiggins, R.J. (1994). *Student-Centered Classroom Assessment.* New York: Merrill.

Tucker, M. (1992, June 17). "A New Social Compact for Mastery in Education." *Education Week: Special Report,* p. S3.

Wood, R. (1977). "Multiple Choice: A State of the Art Report." *Evaluation in Education* 1, 3: 191-280.

# I-7

# Technical Quality Matters

JOAN HERMAN
NATIONAL CENTER FOR RESEARCH ON EVALUATION, STANDARDS,
AND STUDENT TESTING (CRESST)

Alternative assessment holds a lot of promise: Assessments that capture significant, meaningful learning and thinking skills; assessments that will be instructionally useful; assessments that can be integrated with instruction; assessments that will promote both learning and accountability; assessments that help—not hurt—teachers and students. Advocates breathe a sigh of relief that curriculum and instruction at last seem to be in the lead; meaningful standards and learning matter most, not obscure measurement concepts, psychometrics, and bewildering statistical formulas.

But concern for learning and instruction must not quell the need for technical quality in assessment. Face validity—looking interesting, looking authentic, looking instructionally useful, looking like problem solving—is not enough. If we want the new alternatives to provide accurate information for sound decision making then they must function as reliable and valid measures of student accomplishment.

## Quality Issues

### What Are Critical Qualities for Assessment Quality?

Validity is the term the measurement community has used to characterize the quality of an assessment: at the simplest level, if test scores accurately reflect the knowledge, skills, and abilities they intend to measure and if decisions based on test results will be good. For traditional multiple-choice measures, concerns for validity have focused on reliability and patterns of relationships that may suggest whether or not the assessment is tapping the intended construct. For example, does a student's performance on a standardized test of problem solving coincide with classroom observations of that student's capability, with success in subsequent courses emphasizing problem solving, and with future life success in handling complex problems?

The notion that students' performance represents something of importance, something beyond the specific task assessed is essential—that is, that test performance, whatever type the measure, generalizes to a larger domain of knowledge and skill. For example, when an assessment asks a student to conduct a hands-on experiment to determine the optimal environment for a silkworm, we probably want to use the student's performance on this specific task to see if the student can use the scientific method to solve problems. We expect the test to represent something more than the specific object included on the assessment.

Admittedly, past indices of reliability and validity have been unduly narrow, and tests that have met traditional quality standards have sometimes been misused, but the underlying concept of validity remains crucial for any assessment. We need an expanded set of criteria for judging validity, building on both the trials and successes of assessment research and policy over the last decade. Researchers at the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) have proposed such a set (Baker 1993; Linn, Baker, and Dunbar 1991) described here.

**Consequences.** The history of testing has many examples of good intentions gone awry. Consequences of an assessment influence how people respond to its results and can influence the validity of those results. This overarching criterion requires planning from the outset to assess the actual consequences of an assessment policy.

**Fairness.** Does the assessment fairly consider students' cultural background? Have all students had equal opportunity to learn the complex thinking and problem-solving skills that are the targets of the assessments? What other evidence is needed to ensure that students have equitable opportunity to succeed?

**Transfer and Generalizability.** This criterion asks if assessment results support accurate generalizations about student capability. Are the results reliable across raters, topics, and subject matters? Consistent in meaning across administrative procedures and locations? This criterion also raises issues of task comparability: How can we know that assessments from different states or locales equitably represent the same standard? Or from an individual student perspective, can we find ways to honor diversity and to provide students with truly alternate ways to exhibit the same high standards?

**Cognitive Complexity.** The requirement that students demonstrate complex thinking skills is a prime rationale for new types of assessment. Yet, we cannot tell by simply looking at an assessment what it assesses: Looking interesting and challenging is not sufficient. How do we know if a performance assessment calls on students' problem-solving skills? How do we know the level of cognitive demand it places on students? We need new methodologies to answer such questions. These answers also are critical in addressing comparability issues raised by current proposals for national standards and a national system of tests. We must create assessments of equivalent cognitive complexity and verify that alternative tasks tap equivalent levels of cognitive challenge.

**Content Quality.** Tasks that measure a given content domain should be worthy of the time and efforts of students and raters.

**Linguistic Appropriateness.** All students must receive credit for what they know and can do, therefore, we need to understand the role of language in the creation of alternative but equally challenging measurement procedures.

**Instructional Sensitivity.** An interesting and challenging assessment task does not always tap knowledge and skills learned in school. Does optimal instruction produce improvements in performance—or does an assessment tap individual differences immune to school efforts? Accurate inferences about the quality of schools or about student progress demand evidence that performance assessments are sensitive to instructional effort. Yet such evidence has not been well developed.

**Curricular Importance.** How important are goals measured by performance assessments? Are these goals valued, powerful, and deserving of the time and energy to teach and measure them? Just as multiple-choice tests may have negatively influenced performance, we must be sure the novelty of performance assessments does not overshadow the underlying importance of learning to apply subject matter in complex situations.

**Content Coverage.** This area raises issues of curriculum match and if the assessment tasks represent the breadth of content to be learned. Because time constraints limit the number of alternative assessments that can be given, adequate content coverage represents a significant challenge.

**Meaningfulness.** One rational for more contextualized assessments is that they ensure students are engaged in meaningful educational experiences that result in greater understanding of the task and stronger motivation for performance. However, additional evidence is needed to support this theory.

**Practicality and Cost.** With more labor-intensive, performance-based assessments, greater attention must be given to logistics, efficient data collection designs, and scoring procedures, as well as to the combinations of assessments that promote validity, utility, and feasibility. Costs need to be judged relative to evidence of benefit, implying a link back to the consequences criterion.

## How Far Along Are We in Ensuring Such Quality?

Empirical data about the quality of new alternative assessments or about their integrity as measures of significant student learning are scant. Research efforts, however, are underway.

### Ensuring Reliable Scoring

One area of both strength and challenge relates to issues of transfer and generalizability. We have accumulated considerable knowledge about reliably scoring essays and other open-ended responses (Baker 1991). We know how to train raters to consistently score open-ended items, and we know how to maintain rater agreement throughout a rating period, across years, and across rating groups. We also know that systematic rater training reduces the number of required ratings and the costs of large-scale assessment (Baker 1991, p. 3). At the same time, achieving such consistency for new, complex assessments is easier said than done. Vermont's experience in scoring its writing and mathematics portfolios, for example, suggests that reaching consensus on desired qualities to be rated and ensuring adequate training can be a problem for large-scale ventures (Koretz et al. 1993).

What is required for effective training? A well-tuned scoring scheme, examples showing the full range of student work likely to be encountered, benchmarks for each score point, a lot of practice in applying the scale to diverse examples of student work, and qualifying checks before raters begin scoring (Herman, Aschbacher, and Winters 1992).

### The Meaning of These Scores

What the scores represent—no matter how reliable the scoring process—remains a vexing issue. Shavelson and his colleagues examined results from hands-on assessments in math and science (Shavelson, Baxter, and Pine 1990; Shavelson, Gao, and Baxter 1991; Shavelson et al. 1990) to answer the question, "How many tasks does one need to get a stable estimate of a student's problem-solving capability in a given topic area?" Their analyses found that about 8 to 20 tasks were needed to obtain reliable individual level estimates. They also estimated that at least 10 different topic areas may be needed to provide dependable measures of student accomplishment in a general subject area. Subsequent work by both Shavelson, Gao, and Baxter (1993) and Baker, Niemi, and Herl (in press) suggest that well-specified domains reduce the number of items needed, but the numbers across domains and subjects still amount to a substantial sum. Given the time required for administering a single hands-on experiment or an extended problem-solving task, these findings give pause for thought.

Shavelson and others have also found that the context in which students perform influences the results. For example, Shavelson looked at how students' perform on science experiments compared with that on simulations and that on journals, all intended to measure the same aspects of problem solving (Shavelson et al. 1993). Similarly, Herman, Gearhart, and Baker (1993), in a study of portfolio assessment, compared how students' writing was judged when based on their writing portfolios, their classroom narrative assignments, and their responses to a standard narrative prompt. Both studies showed substantial variation in individual performance across the assessment tasks. What students do and the circumstances under which they do it influences both their performance and inferences about their capabilities.

### Ensuring Fairness

Among the strengths of alternative assessments are their attention to process and their potential for integration with instruction. While these characteristics may enable some students to better show their capabilities, these same characteristics may overestimate performance for other students. Consider that students often get support in planning, drafting, and revising work done as part of a classroom assignment; in fact, this support is a hallmark of good instruction. But at the same time, does this additional support constitute learning for an individual student or does it simply make an individual student's work look better? When used for classroom assessment purposes only, this question is not a major issue. Teachers, after all, have many indicators of student capability and know the conditions under which work is produced. The issue becomes important in large-scale assessment settings where the question is "What can an *individual* do?"

Fairness is an issue here as well. If some students get more help than others, are they unfairly advantaged during the assessment? A study by Gearhart et al. (1993) in a small sample of elementary school classes indicates that teachers differ in the amount of support they give to individual students. When students have different levels of assistance, how do we assess their work to determine what they actually can do individually? And how do we provide equitable assessment settings?

The popularity of group assessment raises additional complications for the "Whose work is it?" question. Webb (1993) found substantial differences in students' performance between cooperative group work and individual work. Low-ability students had higher scores based on group work than on individual work: Students could not necessarily accomplish alone what they were able to accomplish in groups. Further, group products don't tell us about the capabilities of individual members.

Still other complications arise when class work merges with homework. In this case, the amount of help students get from family and friends becomes an additional threat to the validity of interpretations about individual student scores. Compare a student whose overzealous, screenwriter parent helps to embellish a composition compared with the student who receives no assistance whatsoever. Or consider the school in a community of highly educated professionals, involved in their children's schooling and able to give a lot of help with homework. Will extended alternative assessments unfairly make these individual students or these advantaged schools look better than those where students receive substantially less help? Differential support issues become important when student work is scored remotely and scores are used to make high-stakes decisions about students or schools.

In any event, and despite the claims of some advocates that alternative assessments will be fairer assessments of students' accomplishments than traditional measures, the equity of portfolio assessment deserves continuing scrutiny. Research to date suggests that patterns of performance on alternative assessments mirror those on traditional measures in terms of relative performance levels of economically disadvantaged minority groups (LeMahieu, Gitomer, and Eresh 1993).

### Emerging Findings Regarding Other Validity Criteria

Concerns for linguistic appropriateness, based on available evidence, are well founded. On the one hand, preliminary findings from Abedi (1994) clearly indicate that students whose first language is not English suffer when items are linguistically complex, and Baker's results indicate that some students perform significantly better when they can use alternatives to essays to represent their knowledge (Baker et al., in press).

Meanwhile, anecdotal evidence meanwhile suggests that content quality, curriculum importance, and content coverage are difficult to balance. Observers have noted that in the quest to find authentic thinking or problem-solving tasks, disciplinary content sometimes get short shrift. Dennie Wolf and Joan Baron, for example, are both concerned that key concepts and principles within the disciplines are not adequately represented by "authentic" assessments.

### Consequences—Good News at a Price?

Evidence about the effects of alternative assessment on curriculum and instruction is weak, but provocative. The great majority of teachers queried in Vermont, for example, reported that the implementation of mathematics portfolios led them to devote more classroom time to teaching problem-solving strategies, and about half indicated they spent more time helping their students deal with mathematical patterns and relationships (Koretz et al. 1993). More than two-thirds of the teachers noted an increase in emphasis on mathematical communication, and approximately half reported engaging students in more small group work. The majority of principals interviewed affirmed that Vermont's portfolio assessment program had beneficial effects on curriculum and instruction, citing as examples specific changes in curriculum content and instructional strategies. More than half the principals suggested that portfolios had value as an educational intervention to promote change.

Similarly, Aschbacher's action research suggests that involvement in the development and implementation of alternative assessments influences teachers' instructional practices and their attitudes toward students (Aschbacher 1992). Two-thirds of the teachers in her study, who received training and follow-up technical support in assessment development and scoring, reported substantial change in the ways they thought about their teaching. At the same time, teachers worry that they can't teach in these new ways and still adequately cover their required curriculum (Borko and Harvey 1993; Davinroy and Hiebert 1993; Flexer, Thomas, and Ferguson 1993; Shepard 1993).

While the research is promising regarding potentially positive effects of alternative assessment on curriculum and instruction, it also indicates substantial time and implementation challenges. Studies consistently report substantial demands on teachers' time (Aschbacher 1992, Koretz et al. 1993, Wolf and Gearhart 1993): Time for teachers to learn new assessment practices, to understand what should be in portfolios and how to help students compile them, to develop portfolio tasks, to discern and apply criteria for assessing students' work, to fine-tune their instructional and assessment practices, and to manage the logistics.

The costs of large-scale performance assessment programs are many: professional development; development of task specifications and prompts; and administration, including observing, recording and compiling documentation, scoring, reporting, revising, and validating (Catterall and Winters, in press). But all of these are needed.

## The Complex Challenge of Ensuring Quality

The development and validation of new kinds of assessments offer great potential and significant challenge. The promise is alluring, yet what we know about alternative or performance-based measures is small when compared with what we have yet to discover. Assessment practices must be accountable to criteria that define quality assessments. These criteria force attention to traditional technical issues, to the consequences of assessment, and to students' opportunity to learn what is assessed. Below is a checklist to help ensure the quality and benefits of new developments.

**Consequences:**

- Is the assessment worth the instructional time?
- Does the assessment encourage good instruction?

- Does the assessment exemplify valued student outcomes?
- Does the assessment discourage a narrow curricular focus?
- Does the assessment provide ample time for students to complete so that results reflect capability rather than test-taking skill?

**Fairness:**

- Does the assessment tap the knowledge and skills students have had an adequate opportunity to acquire during classroom instruction?
- Is the assessment free of cultural, ethnic, and gender stereotypes?
- Is the assessment free of tasks or situations more familiar to students of one background or gender than to another?
- Does the assessment use a scoring process applied without bias?
- Is the assessment free of difficulties irrelevant to the knowledge or skills being assessed; for example, does the assessment use unnecessarily difficult vocabulary or linguistic structures when assessing science performance?
- Does the assessment enable all students to demonstrate what they know and can do in the areas being assessed ?
- Is the assessment teachable?

**Transfer and Generalizability:**

- Does the assessment describe the domain it intends to assess?
- Does the assessment represent the intended domain; for example, does the assessment include a range of important, significant instances of targeted knowledge and skills?
- Does the assessment provide evidence that results are generalizable—are indicative of student performance in a broader domain of knowledge?
- Does the assessment consider the number of tasks a student must complete in order to yield generalizable results?
- Does the assessment includes explicit criteria for scoring, and preferably a guide describing the application of those criteria ?
- Does the assessment provide evidence that results are consistent across raters and across scoring occasions?

**Cognitive Complexity:**

- Does the assessment use tasks for which students can be expected to have adequate background knowledge?
- For tasks assessing thinking and problem-solving skills, does the assessment pose problems for which students have no prior exposure to the specific tasks used in the assessment?
- Does the assessment use tasks whose solutions cannot be memorized in advance?
- Does the assessment provide evidence that tasks elicit complex understanding or problem-solving skills?

**Content Quality and Coverage:**

- Does the assessment assess key concepts and principles from the disciplines?
- Does the assessment avoid irrelevant information?
- Does the assessment use tasks consistent with current educational pedagogy?
- Has the assessment been reviewed by content experts to ensure quality and disciplinary/interdisciplinary appropriateness of tasks?
- Does the assessment cover a broad range of skills and content?
- Do tasks reflect a rich, balanced curriculum?

**Meaningfulness:**

- Does the assessment evoke positive reactions from both students and teachers?
- Is the assessment credible to students, teachers, parents, and the public as a valid indicator of student competence in the particular assessment area?
- Does the assessment engage and motivate students to do their best?

**Cost and Efficiency:**

- Is the assessment administratively feasible?
- Is the assessment optimally cost-efficient?
- Can the assessment be scored and administered efficiently and effectively?

### References

Abedi, J. (January 1994). "Progress Report on Language Background Project" (meeting minutes). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Aschbacher, P.R. (1992). *Issues in Innovative Assessment for Classroom Practice: Barriers and Facilitators* (final deliverable). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Baker, E.L. (April 1991). "What Probably Works in Alternative Assessment." In *Authentic Assessment: The Rhetoric and the Reality*. Symposium conducted at the annual meeting of the American Educational Research Association, Chicago.

Baker, E.L. (1993). *Continuation Application for OERI National Center for Research on Evaluation, Standards, and Student Testing*. Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Baker, E.L., D. Niemi, and H. Herl. (in press). "Using HyperCard Technology to Measure Understanding." In *Tech-*

*nology Assessment in Education and Training*, edited by E.L. Baker and H.F. O'Neil Jr. Hillsdale, N.J.: Lawerence Erlbaum Associates.

Borko, H., and J. Harvey. (September 1993). "Teachers Putting Performance Assessment Into Practice." Presentation at the annual CRESST (UCLA) Conference, Los Angeles.

Catterall, J., and L. Winters. (1994). *Economic Analysis of Tests: Competency, Certification, and Authentic Assessments* (Tech Rep. No. 373). Los Angeles: University of California, Center for the Study of Evaluation.

Davinroy, K.A.H., and E.H. Hiebert. (December 1993). "An Examination of Teachers' Thinking About Assessment of Expository Text." Paper presented at the annual meeting of the National Research Council, Charleston, S.C.

Flexer, R., A. Thomas, and S. Ferguson. (October 1993). "Performance Assessment in the Classroom." Presentation at the regional meeting of the National Council of Teachers of Mathematics, Durango, Colo.

Gearhart, M., J.L. Herman, E.L. Baker, and A. Whittaker. (1993). *Whose Work Is It? A Question for the Validity of Large-Scale Portfolio Assessment* (Tech. Rep. No. 363). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing; Center for the Study of Evaluation.

Herman, J.L., P.R. Aschbacher, and L. Winters. (1992). *A Practical Guide to Alternative Assessment*. Alexandria, Va.: Association for Curriculum and Supervision Development.

Herman, J.L., M. Gearhart, and E.L. Baker. (Summer 1993). "Assessing Writing Portfolios: Issues in the Validity and Meaning of Scores." *Educational Assessment* 1, 3: 201-224.

Herman, J.L., D.C.D. Klein, T.M. Heath, and S.T. Wakai. (1994). "Preliminary Results on Equity in Alternative Assessment" (draft deliverable). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Koretz, D., B. Stecher, S. Klein; D. McCaffrey, and E. Deibert. (1993). *Can Portfolios Assess Student Performance and Influence Instruction? The 1991-92 Vermont Experience* (Tech. Rep. No. 371). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing (CRESST).

LeMahieu, P., D.H. Gitomer, and J.T. Eresh. (1993). "Portfolios in Large-Scale Assessment: Difficult but Not Impossible." Unpublished manuscript, University of Delaware.

Linn, R., E. Baker, and S. Dunbar. (1991). "Complex, Performance-Based Assessment: Expectations and Validation Criteria." *Educational Researcher* 20, 8: 15-21.

Shavelson, R., G.P. Baxter, and J. Pine. (October 1990). "What Alternative Assessments Look Like in Science." Paper presented at the Office of Educational Research and Improvement Conference, "The Promise and Peril of Alternative Assessment," Washington, D.C.

Shavelson, R., X. Gao, and G. Baxter. (November 1991). *Design Theory and Psychometrics for Complex Performance Assessment: Transfer and Generalizability* (Project 2.4 Progress Report, November 30, 1991, to OERI, Grant no. R117G10027). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Shavelson, R., X. Gao, G. Baxter. (1993). *Sampling Variability of Performance Assessments* (Tech. Rep. No. 361). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Shavelson, R., P. Mayberry, W. Li, and N.M. Webb. (1990). "Generalizability of Job Performance Measurements: Marine Corps Rifelman." *Military Psychology* 2, 3: 129-144.

Shepard, L. (Ed.). (1993). *Report on Case Study of the Effects of Alternative Assessment in Instruction, Student Learning and Accountability Practices* (final deliverable). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Webb, N.M. (1993). *Collaborative Group Versus Individual Assessment in Mathematics: Group Processes and Outcomes* (Tech. Rep. No. 352). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Wolf, S., and M. Gearhart. (1993). *Writing What You Read: Assessment as a Learning Event* (Tech. Rep. No. 358). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

# 1-8

# Authentic Assessment in the Classroom: Principles and Practices

PETER WINOGRAD
UNIVERSITY OF KENTUCKY, LEXINGTON

FRAN DAVIS PERKINS
UNIVERSITY OF ALABAMA, TUSCALOOSA

## The Need for Authentic Assessment

In her storybook *First Grade Takes a Test,* Miriam Cohen (1980) makes poignantly clear the concerns regarding traditional methods of assessment. The story is told from the perspective of a 1st grade teacher and her students, and their words mesh beautifully with Lillian Hoban's illustrations to show us how students begin to question their self-worth and competence when they find out that only one student in the class "did a good test." Fortunately, the students' teacher, like most good teachers, understands the strengths of her students and their need to feel good about themselves as learners, and she responds appropriately: "The test doesn't tell us everything. It doesn't tell us all the things you can do! You can build things! You can read books! You can make pictures! You have good ideas! . . . Those are important things."

This classroom scenario is not unique. Many teachers across the country are acutely aware that tests don't tell us everything. In fact, they know that no single method of assessment can provide all the information we need. They also recognize that traditional methods of assessment (i.e., standardized tests, criterion-referenced tests, teacher-made tests, and commercially developed tests) often ignore the instructional contexts in which the learning occurred and are therefore outdated and inappropriate models of assessment (Neill and Medina 1989; Shepard 1989; Winograd, Paris, and Bridge 1991). Yet many feel that traditional testing will remain a way of life in classrooms because of the use (or misuse) of tests in judging the effectiveness of schools and teachers, in placing students in special programs, in retaining or passing students, in making policy and curricular decisions, and in determining whether programs receive funding (Routman 1991; Tierney, Carter, and Desai 1991).

The recent period of educational reform, however, has proven rich in possibilities for those interested in using assessment in more constructive ways to support children's learning. In this chapter, we identify some of the principles underlying the move to more constructive assessments, and we examine specific techniques and strategies that teachers can use in their classrooms. Our experiences are concerned primarily with young children learning to be literate, so most of our examples are drawn from reading and writing instruction. Authentic assessment, however, is relevant throughout the grades and across the curriculum.

## What Is Authentic Assessment?

Authentic assessment is assessment that occurs continually in the context of a meaningful learning environment and reflects actual and worthwhile learning experiences that can be documented through observation, anecdotal records, journals, logs, work samples, conferences, portfolios, writing, discussions, experiments, presentations, exhibits, projects, and other methods. Authentic assessments may include individual as well as group tasks. The emphasis is on self-reflection, understanding, and growth rather than on responses based only on the recall of isolated facts.

The intent of authentic assessment is to involve learners in tasks that require them to apply knowledge in real-world experiences. Authentic assessment discourages paper-and-pencil tests that are disconnected from the teaching and learning actually taking place. In authentic assessment, there is a personal intent, a reason to engage, and a genuine audience beyond the teacher (Crafton 1991).

Imagine a 5th grade classroom in which students collaborate to create a school newspaper for the purpose of sharing school and community news. The paper will include editorials, recipes, feature stories, and consumer reports. Imagine also that students help develop criteria for assessing their individual and class projects. This project provides students with opportunities to demonstrate their ability to organize ideas, write, research, advertise, and communicate effectively with the public. It allows them to receive instruction in the various types of writing necessary to complete the project, and to make adjustments and revise content as they reflect on the learning process with their teacher and classmates until they "go to press." These real-life experiences remind us "that school is not for getting ready to do the real stuff of life sometime in the distant future, it is for doing real things, for real audiences, and for ourselves, right now" (Crafton, p. 13). This project illustrates the interactions between assessment and instruction that are undetectable through traditional methods of evaluation. In this example, assessment not only informed instruction, it helped children to become actively involved in important processes.

## Principles of Authentic Classroom Assessment

How do we address authentic assessment? We find it helpful to keep four principles in mind when we think about authentic assessment.

### Principle #1: Authentic assessment focuses on meaningful goals.

The first question is "What do we want to measure?" Identifying what is to be measured, however, is often a difficult task. Goals for students can be placed on a continuum ranging from the very broad and general to the extremely specific and narrow. One of the 1992 national education goals, for example, states that students will demonstrate "competency in challenging subject matter, including English, mathematics, science, history, and geography; and every school in America will ensure that all students learn to use their minds well, so they may be prepared for responsible citizenship, further learning, and productive employment in our modern economy." This is obviously an important goal, but it is too broad to measure in any manageable way. At the other end of the continuum are very low-level goals that focus on isolated skills: "Students should identify the sound the letter *a* represents in the word *man*." Specific goals like this one are often easy to measure, but they may not be worth the effort.

Developing goals at just the right level on the continuum requires teachers to think about meaningful, authentic outcomes that are observable and useful to those who use the information gained from the assessment. For example, "Students will become more self-reflective about their reading and writing" might be a useful classroom goal for teachers interested in helping students become more independent learners. "Students are able to understand a wide variety of authentic reading materials" might be a useful goal for teachers who want to focus on comprehension. And "Students will develop an interest in reading" might be a useful goal for teachers interested in helping students become lifelong readers.

Even after identifying meaningful goals, teachers must still ensure that the goals actually drive assessment and instruction. What happens too often is that meaningful goals are included in the language of curriculum documents, but the district or school still relies heavily on textbooks and traditional multiple-choice tests. The teachers, then, end up with the textbook and test maker's scope and sequence as their de facto goals of assessment and instruction. Although the curriculum may identify, for example, helping students develop a lifelong love of reading as an important goal, students are evaluated only on their ability to read short paragraphs and select the right answer from a list of choices.

### Principle #2: Authentic assessment enhances learning rather than simply serving the purpose of accountability.

One of the most important lessons to learn about assessment is that it can and should serve different purposes (Winograd, Paris, and Bridge 1991). We find it useful to think about the various purposes of assessment in terms of different audiences. Here are some examples of how student assessments can affect different audiences:

- Students become more self-reflective and in control of their own learning.
- Teachers focus their instruction more effectively.
- Educators determine which students are eligible for Chapter 1, programs for the gifted, or special education.
- Parents understand more about their own children's progress as learners.
- Administrators understand how groups of students in their schools are progressing as learners.
- Legislators and other representatives of the public understand how groups of students across the state are progressing as learners.

Each of these audiences, from students to legislators, is important and worthy of the best information possible.

In the past, however, assessment for students has been neglected in favor of assessment for accountability. The primary purpose of authentic assessment is to nurture children's growth in understanding, self-reliance, and confidence (Johnston 1991). Thus, one of the important changes in the way we think about testing is that students are a valuable source of information for teachers as they plan their instruction. Students are also a valid and valuable source of information about a teacher's instruction. We have often heard students say to their teachers, "I don't understand what you mean!" or "That doesn't make sense to me!" or "I thought we did a great presentation this time!" or "This is the best reading I have ever done!" These informal and unplanned remarks inform the teacher about the lesson—its appropriateness, its relevance, or the need to give students further assistance.

We want to stress that assessment for learning and assessment for accountability are not mutually exclusive. In fact, one of the most important aspects of the current wave of educational reform is that assessment for learning and assessment for accountability are aligned. This is an important advance in the way we think about assessment because it assumes that the more we help students become self-reflective and take control of their learning, and the more we help teachers improve their instruction, then the better students will perform on tests of accountability. When properly aligned, assessment for learning and assessment for accountability reinforce one another and focus on similar goals.

### Principle #3: Authentic assessment empowers teachers by helping them build the expertise necessary to trust their own professional judgments about learners.

For many years, teachers have made on-the-spot decisions during instruction based on student responses. By observing, listening, and responding to their students, teachers make judgments about whether to continue with a specific type of instruction, adjust their teaching strategy for the lesson, or abandon the lesson until students have built a greater knowledge base. A test after the lesson doesn't allow teachers to make such immediate judgments. We encourage teachers to trust such judgments, but we stress the importance of using authentic assessments as a way of collecting evidence to back up those judgments.

Teachers can also use their professional judgment to consider what types of evidence can form the basis for assessment. As Wiggins (1989, p. 705) asserts, "Do we judge our students to be deficient in writing, speaking, listening, artistic creation, finding and citing evidence, and problem solving? Then let the tests ask them to write, speak, listen, create, do original research, and solve problems."

### Principle #4: Authentic assessment provides teachers with richer opportunities to engage in linguistically and culturally appropriate evaluation and instruction.

Cultural diversity must be a consideration when planning for assessment and instruction. Far too often the voices of all children are not heard or validated in instructional activities or assessment. The reasons are rooted in misconceptions and stereotypes about linguistically and culturally diverse learners. King (1993, p. 174) maintains that teachers must commit to "finding out about their students' way of viewing so that they can better understand them and so that they can help their students expand their perceptions and views of the world to include those of other cultures, including that of the larger general society."

Children's language and diverse experiences must be accepted and used as a basis for assessment. For example, teachers may collect children's personal and family stories, which are naturally part of their culture. These stories may be used by the teacher to evaluate a child's awareness of story structure or as a source of information for future instruction.

King also suggests that teachers should consider patterns of social interaction, diverse learning styles, and the ways people use oral and written language across various cultures. Such considerations would have a direct impact on the way teachers assess and instruct learners. Ultimately, teachers must ask themselves if their assessment and instructional practices appreciate and accept the linguistic and cultural diversity of all learners.

## Classroom Strategies for Authentic Assessment

In the remainder of this chapter we examine four authentic assessment strategies: portfolios, observation (with anecdotal records and developmental checklists), student-teacher conferences, and report cards. These by no means exhaust all the possibilities. It is important to note from the beginning that the four strategies are interrelated. The portfolios could contain, among other things, records of the teacher's observations and conferences with students. Portfolios, anecdotal records, and developmental checklists can provide important starting points for discussions among teachers, students, and parents.

### Portfolios

Portfolios come in a variety of forms, and each serves a different purpose. One of the most popular types of portfolios is a *best-piece portfolio,* which contains examples of coursework that a student (sometimes with help from the teacher) considers to be of the highest quality.

Another kind of portfolio is a *descriptive portfolio,* which assembles a wide variety of measures on children. For example, teachers in the South Brunswick Public Schools in New Jersey use a descriptive portfolio to help them better document progress for children in grades K–2. These portfolios include self-portraits drawn by the child; interviews with the child and with the parents; the results from the Concepts About Print test; a word awareness writing activity; a reading sample; and a writing sample.

Many teachers are familiar with *process portfolios,* which are an intrinsic part of the writing process advocated by Graves (1983) and others. Process portfolios can contain a variety of products, including finished projects, manuscripts in progress, possible topics for future papers, and lists identifying ways in which the student has grown as a reader and a writer.

Finally, a number of states are requiring teachers to gather *accountability portfolios.* Kentucky, for example, requires teachers and students in the 4th, 8th, and 12th grades to assemble portfolios that include a table of contents; a best piece of work; a letter to the reviewer telling why the piece was selected as one of the best; a short story, poem or play; a personal narrative; as well as other samples of writing. This portfolio, along with samples of writing done on demand, provides part of the evidence by which the state judges the quality of instruction.

Best-piece portfolios, descriptive portfolios, process portfolios, and accountability portfolios are not the only kinds of portfolios in use. Nor are they pure types that are clearly distinct from one another. Teachers and students often find that the portfolios they use in class may serve different functions at different times. In actual use, it is often hard to categorize portfolios as purely best-piece, descriptive, or process portfolios.

Portfolios are most effective when they encourage students to become more reflective about and involved in their own learning. Indeed, a number of writers argue that what makes a portfolio a portfolio is that these collections include "student participation in selecting contents, the criteria for selection, the criteria for judging merit, and evidence of student self-evaluation" (Paulson, Paulson, and Meyer 1991, p. 60). Tierney, Carter, and Desai (1991) also stress that portfolios are not objects, but rather represent the students' ability to engage in the processes of selecting, comparing, self-evaluation, sharing, and goal setting.

Helping children reflect on their own work forms the basis for a rich interplay between assessment and instruction. For example, one experienced teacher, Gretchyn Turpin, had her class come up with ideas about what they thought makes a piece of writing a "best piece." Her students generated the following list:

- Uses imagination.
- The words make good pictures in your mind.
- Uses things that happen in your own life.
- Uses writing rules like periods and capitals.
- Spaces between words.
- Makes you want to keep on reading.

"After writing down these ideas," Turpin told us, "the children were able to evaluate the work in their portfolios with some focus. The audience for this best piece was their own classmates. When they realized that their best pieces would truly circulate among their friends, they became very self-motivated and diligent in their efforts to present their best pieces to their friends in the best possible light. Now when someone reads their best piece in the classroom, the author isn't satisfied unless he or she is getting laughs, gasps, or smiles from the audience. We now choose a best piece about once every two months, which has added a lot of enthusiasm to our writing workshops."

What the students were experiencing in Turpin's class was the essence of assessment for learning rather than assessment for accountability. The opportunity to reflect on their own work leads students to better writing, which in turn leads to more sophisticated reflections on their own work. Students who have this kind of opportunity learn the most important secrets to becoming a good writer.

### Observation: Anecdotal Records and Developmental Checklists

Most of the information that teachers gather about their students comes from observation. Observation can take many forms, from informal observation on the playground to systematic anecdotal records to more structured observations using checklists.

Anecdotal records, developmental checklists, and other forms of observation offer many benefits:

- They provide teachers a way to assess how students interact with a complex environment both in and out of classrooms. Teachers gain a more valid understanding of what each student really knows and can do.
- They provide teachers an efficient method for assessing students in many different situations over long periods of time, thus increasing the reliability of the assessment data.
- They focus the teacher's attention on what the student can do, rather than on what the student has yet to learn.
- They provide a relatively stress-free form of evaluation for students, especially those students who become anxious when they take standardized tests.

*Anecdotal records* are informal observations about what students are learning, how students are responding to instruction, or any other student behaviors, actions, or reactions that might provide teachers with some insight (Rhodes and Nathenson-Mejia 1992). Anecdotal records often focus on questions such as these:

- What can this child do?
- What does this child know?
- How does this child read, write, work on projects, work with others, or deal with other important aspects of the school curriculum?
- What other kinds of questions does the child have about his or her work?
- What does the child's attitude reveal about his or her growth and progress?

In addition, anecdotal records often include teacher comments and questions that are particularly useful in helping teachers become more reflective about their teaching. Yvonne Siu-Runyan (1991) offers a powerful example of how anecdotal records can help teachers improve their instruction:

> 8/31/87: Most of the children are struggling with their pieces. They want to write grand stories, but don't know enough about their topics to write well. They want to write fiction, and not personal narratives which is what I think would make a difference to them.
>
> 9/2/87: The students are abandoning pieces like mad. Even though I did a lesson on brainstorming ideas for writing topics, most of them are still struggling with topic choice.
>
> 9/14/87: I wish the kids were more interested in writing. They are wasting a lot of time because many don't know what to write about. I am getting worried.
>
> 10/7/87: Kids of concern: Zephyr, Tom, Travis, Geoff, Cara, Ashleigh, Lucas, John, Ben, Missy, Mel, Rojana, and Mike. These kids are still not interested in their writing. I'm beginning to feel like a failure.

When Siu-Runyan reviewed her anecdotal records, she knew that she needed to respond to these children quickly. She also realized that she had not written with the children or modeled how literature helped her own writing. She decided to try modeling the kinds of writing she wanted her students to learn. The following anecdotal records reveal what happened after she spent time modeling writing in response to literature:

> 10/1/87: Lucas and Melissa are writing their own important stories. I am so pleased that they are trying their hand at this. Perhaps now others will use literature to help them learn about writing. Now to keep the momentum going. I must be careful about what books I choose to share. I must remember to select personal narratives with an obvious design. I think this will help the kids a lot.
>
> 10/20/87: Read *My Mom Travels A Lot* (1981) by Caroline Feller Bauer. Discussed the design of the book. Created several stories verbally following the opposite design of this book. Today Cara developed her own story from the model presented. Great!

*A developmental continuum* is another common and useful tool for observing students' progress in the primary program. A major reason for its appeal is that it can be used in a wide variety of situations. Figure 1 shows an example of such a continuum. Teachers interested in looking at other examples of a developmental continuum will find lots of good ideas in the literature on assessment (e.g., Clay 1985; Harp 1991; Kemp 1989; Routman 1988, 1991; Sharp 1989).

FIGURE 1

## SAMPLE DEVELOPMENTAL CONTINUUM

Student name ______________________________ Date ____________

### Juneau Primary Writing Rubric

| | IDEAS AND CONTENT | CONVENTIONS |
|---|---|---|
| 1 | • **Communicates meaning through (single) picture(s).**<br>• Draws/scribbles as writing.<br>• Writes random letters or letter strings. | • **Writes shapes that look like letters.** |
| 2 | • **Communicates meaning with pictures supported by a word/label.**<br>and/or<br>• Communicates a story through a related series of pictures.<br>• Writes names of people/things or environmental print. | • **Uses mostly uppercase letters.** |
| 3 | • **Begins to communicate meaning with words.**<br>• Communicates meaning to reader through connected words. | • **Writes some recognizable, understandable words.**<br>• Attempts phonetic, temporary/invented spelling (using consonant beginning and/or ending sounds).<br>• Writes left to right, top to bottom.<br>• Begins to use spaces between words. |
| 4 | • **Begins to write simple sentences, expressing a complete thought.**<br>• Ideas may be expressed in lists or patterned statements. | • **Writes simple sentences or pattern sentences.**<br>• May use repetitive sentence structure.<br>• Appropriately uses uppercase and lowercase letters.<br>• Begins to use capitals and some punctuation correctly.<br>• Uses vowels in temporary/invented spelling.<br>• Begins to use conventional spelling.<br>• Consistently uses spaces between words. |
| 5 | • **Begins to develop story sense.**<br>• Writer's emotion/enthusiasm begins to show.<br>• Displays a sense of beginning and ending.<br>• Sequence of ideas is logical.<br>• Uses specific, accurate, suitable vocabulary. | • **Begins to use a variety of sentence structures.**<br>• Generally uses capitals and periods correctly.<br>• Frequently uses conventional spelling on high-frequency words. |
| 6 | • **Begins to add supportive details.**<br>• Generally stays on topic.<br>• Writer's emotion/enthusiasm shows.<br>• Occasionally uses creative or vivid words/phrases appropriately.<br>• Story has a beginning and an end. | • **Uses a wide variety of sentence structures.**<br>• Begins to punctuate correctly.<br>• Consistently uses conventional spelling on most words. |
| 7 | • **Write in a convincing, natural style.**<br>• Represents one main, focused idea/topic.<br>• Uses detail(s) to develop idea.<br>• Holds reader's attention.<br>• Organizes story with clear beginning, middle, and end.<br>• Communicates feelings to reader.<br>• Paints a picture in reader's mind.<br>• Experiments with uncommon words and new and/or different ways of saying things. | • **Uses varied and complex sentence syntax (construction of word, phrases, or clause in a sentence).**<br>• Write clearly—paper is easily understood.<br>• Consistently uses punctuation marks and capitalization correctly.<br>• Consistently uses conventional spelling. |

NOTE: You can use this rubric with some forms of writing, particularly fiction and nonfiction. If a paper contains only the child's name, it should not be factored into conventions. Pictures may be present at any level.

Reprinted with permission of the Juneau School District, 10014 Crazy Horse Drive, Juneau, AK 99801 Revised 8/94

### Conferences

Conferences between teachers and students are another powerful tool in the repertoire of teachers adept at authentic assessment. Conferences can be quick and informal or they can be more structured and systematic. Informal conferences may last from three to four minutes and may focus on something interesting that the teacher [illegible]en or overheard. More structured conferences [illegible] take longer and follow a predictable pattern in [illegible] students have a good idea of what is expected of them. For example, if the conference is over books read or papers written, then students are familiar with the kinds of questions that will be asked. The important point about conferences of all sorts is that they should be conducted in a safe and comfortable manner so that students feel encouraged to take risks and share their ideas.

Conferences can provide information about many areas:

- What the student is learning
- The student's understanding of reading, writing, mathematics, and other aspects of the primary program curriculum
- The student's interests
- Areas in which the student needs help
- What the teacher is doing that the student believes is particularly helpful
- What the student would like to learn next

Here are some examples of different kinds of conferences and the kinds of information that teachers can learn from them:

**Reading Conferences.** The teacher can listen to the student read, discuss a book that has been previously read, or talk about the student's book log. These activities enable a teacher to learn about:

- The strategies the student uses
- Whether the student reads for meaning
- Whether the student is developing fluency
- The student's interests
- Whether the student selects reading materials of appropriate difficulty
- Whether the student is reading a variety of genres
- The student's progress in comprehending and retelling what has been read
- The student's ability to justify an opinion about what has been read

**Writing Conferences.** The teacher can help the student brainstorm topics to write about, discuss early drafts, listen to a student read a paper, or talk about which piece of work the student considers his or her best work. These activities enable teachers to learn:

- How the student is progressing in the use of the writing process
- What the student knows about organization, topic development, mechanics, and spelling
- Whether the student can effectively verbalize opinions, ideas, and feelings
- Whether the student can write for a variety of purposes
- Whether the student can edit drafts to a point where others can understand them

### Report Cards

Changing report cards to accurately communicate a child's progress is one of the most important challenges facing teachers who use authentic assessment. The following suggestions may prove helpful to teachers and principals as they revise their report cards to reflect current assessment practices:

- Include a clear statement of the philosophy and the goals of the primary program.
- Focus on the child's growth in all areas throughout the year.
- Avoid blame and harmful labels.
- Accentuate the developmental nature of learning.
- Address areas of concern (if and when they are identified) in a constructive manner.
- Include space for the teacher to write narrative descriptions.
- Include space for parents to comment and respond.

The Saffell Street School Primary Progress Report, shown in Figure 2, is a creative example of a report card that reflects authentic assessment. It is an exemplary report card for a number of reasons. First, the teachers have identified the important *goals* in their curriculum. For example, they want their students to display an interest in reading and writing; engage in important classroom activities, including Super Silent Reading, group reading, reading discussion, and writing workshop; apply important skills and strategies in reading writing, math, listening, and speaking; and develop appropriate personal and social behaviors.

Second, the teachers at Saffell Street School use *rubrics* to help parents understand the developmental nature of learning. Here, for example, is a rubric that focuses on a student's performance during a reading discussion:

FIGURE 2

## SAFFELL STREET SCHOOL PRIMARY PROGRESS REPORT

Student ______________________________ Teacher ______________________________

### LANGUAGE AND LITERACY

DISPLAYS INTEREST IN PRINT
____ shares books with others
____ chooses to spend time with books
____ asks to be read to
____ listens attentively to books in a group
____ contributes to group discussions
____ checks out books frequently

DURING SUPER SILENT READING

| | | |
|---|---|---|
| Involves self intensely with print | Maintains interest in print with additional support | Has difficulty being involved with print |

DURING GROUP READING

| | | |
|---|---|---|
| Totally involved with reading experiences | Willing to participate with additional support | Unable to stay involved with group activity |

DURING READING DISCUSSIONS

| | | |
|---|---|---|
| Carries on a meaningful conversation about reading | Willing to answer questions about reading material with guidance | Unable to express thoughts or feelings about reading material |

DISPLAYS INTEREST IN WRITING
____ chooses to write or draw
____ wants more writing time
____ wants to share their writing
____ shows interest in others' writing
____ uses available print

DURING WRITING WORKSHOP

| | | |
|---|---|---|
| Shows independence using the writing process | Shows some independence, needing individual guidance to progress | Needs individual guidance to participate |

STORY DEVELOPMENT

| | | |
|---|---|---|
| Develops and organizes ideas with a purpose | Expresses logical and sequential thoughts and ideas; needs to develop a clearer purpose | Expresses ideas; may or may not be logically connected |

KNOWLEDGE OF MECHANICS

| | | |
|---|---|---|
| Able to use punctuation, capitals, grammar, and complete sentences | Shows some use of punctuation, grammar and complete sentences, but not consistently | Needs awareness of punctuation, capitals, grammar and complete sentences |

READING STRATEGIES (e.g., clues, context, background knowledge, prediction, phonics)

| | | |
|---|---|---|
| Uses a variety of reading strategies | Beginning to use one or more reading strategies | Unaware of reading strategies |

STEPS IN READING
____ listening to stories, unaware of print and its function
____ picture reading, describing pictures
____ pretends reading, turning pages, tracking print, rehearsing silently
____ memory reading
____ recognizes some words but not all
____ recognizes patterns in reading
____ reading with understanding, summarizes

(continued on next page)

BEST COPY AVAILABLE

FIGURE 2—*continued*

## SAFFELL STREET SCHOOL PRIMARY PROGRESS REPORT

Student ______________________ Teacher ______________________

### MATH

PROBLEM SOLVING

| | | |
|---|---|---|
| Able to problem solve independently | Can use problem solving with some guidance | Needs assistance to solve problems |

COMMUNICATES MEANING

| | | |
|---|---|---|
| Uses symbols and words to show meaning | Uses symbols and/or words, needs assistance to show meaning | Needs assistance to use symbols and/or words to show meaning |

NUMBER UNDERSTANDING

| | | |
|---|---|---|
| Demonstrates and extends understanding about number meaning | Demonstrates understanding about number meaning | Beginning to show an awareness of number meaning |

### LISTENING AND SPEAKING

FOLLOWS DIRECTIONS

| | | |
|---|---|---|
| Almost always follows directions independently | Working to develop better listening skills to follow through with directions | Needs constant assistance in following directions |

COMMUNICATES IDEAS VERBALLY

| | | |
|---|---|---|
| Willing and able to communicate effectively with others | Willing to express thoughts; working to develop a clearer focus | Needs encouragement to express ideas |

### PERSONAL AND SOCIAL GROWTH

USE OF TIME

| | | |
|---|---|---|
| Is productive and involved | Sometimes needs encouragement to use time productively | Needs assistance to become involved in productive activities |

EFFECTIVE GROUP MEMBER

| | | |
|---|---|---|
| Works well with others | Needs limited assistance to work with others | Has difficulty working with others |

COOPERATES WITH PEERS AND SCHOOL PERSONNEL

| | | |
|---|---|---|
| Shows respect, gets along well with others | Needs encouragement with limited guidance | Needs constant reminding of how to cooperate |

BEST COPY AVAILABLE

Parents! Please write a home goal for your child. It may be something that you will continue to work on during this next reporting period.

Sign and return ______________________ Date ______________

*During reading discussion [the student]:*

| | | |
|---|---|---|
| carries on a meaningful conversation about reading. | is willing to answer questions about reading material with guidance. | is unable to express thoughts or feelings about reading material. |

As you can see, the rubric shows parents the continuum of developmental behaviors that the teacher is looking for as children grow. Beginning students are unable to clearly express their thoughts about what they have read. Developing students can answer questions about what they have read, but teachers still must provide some modeling and support. Independent students can discuss what they have read with ease and confidence.

Abandoning traditional marking systems that label students "satisfactory" or "unsatisfactory" or give them simple letter grades is an important step for educators committed to making real changes in the ways we conceptualize assessment. Terms like "unsatisfactory" or letter grades like "F" suggest that young students are failures. Developmental rubrics, in contrast, suggest that the student still has time to grow. They also give students and parents a clear idea of the school's expectations.

The move from letter grades to scoring systems that communicate the developmental nature of young children's learning is crucial. Although letter grades indicating clear success or failure may sometimes be appropriate for older students and adults (for instance, in evaluating whether a pilot-in-training has the knowledge and skills needed to fly a plane), they are never appropriate for young children. The philosophy underlying authentic assessment holds that it is wrong to characterize young children as failures simply because they lack certain kinds of knowledge or cannot do specific tasks. Rather, we should view children as inexperienced and provide them more support and opportunities to grow.

Here's a straightforward method of developing your own rubrics:

1. Identify a key instructional task or activity in your classroom (e.g., group reading, writing workshops, cooperative groups working on projects, etc.).

2. Think about some specific students in your class who do very well on this task or activity. Write down some descriptive phrases that capture what they do. These phrases should be placed at the left side of the rubric.

3. Now think about some specific students in your class who have yet to learn how to complete the task or participate in the activity. Write down some descriptive phrases that capture what they do (or cannot do). These phrases should be placed at the right side of the rubric.

4. Finally, fill in the middle of the rubric with phrases that describe students who can do the task or activity occasionally or with support.

Another important section of the Saffell Street School Primary Progress Report is the space allowed for narratives written by the teacher. These brief reports help parents gain a more personal view of their child. Here are teachers' comments about students we'll call Tom and Jenny:

> I love having reading conferences with Tom. He has really begun to be able to discuss a book. He comprehends the story well. We discussed character feelings in Fuzzy Rabbit. He is excellent at reading silent to answer content questions. His oral reading is very fluent and improved.
>
> Jenny says she likes looking at books and reading but she doesn't like to read at home. She would be helped a great deal if she would just read 10 minutes a night (read to her). Her goal in reading is to "get books sne could read."

The Saffell Street School Primary Progress Report also includes a place for parents to write a home goal for the student. Here is what Tom's mother wrote:

> Thanks for the report. We'll work with Tom on speaking in front of people and telling stories.

The opportunity for parent responses provides teachers with some important insights about the children and their families. Consider these examples from different parents:

> My goal for Linda is to have more individual time with her, to have her read to me more & to let her help me cook more (she likes that). She & I made banana bread a few weeks ago. She helped me follow the recipe, get ingredients prepared, & clean up. We had a good time.
>
> I don't know what's happening but Alex is not wanting to go to school now, he says he is bored. He has been complaining for about the last 2 weeks. I told him to talk to you but of course he won't. I don't know what to do.
>
> I let Bill know that I am very proud of him. He has done very good. My home goal is to help Bill with his behavior and working with others. Thank you.

## A Comfortable Fit

Teachers interested in finding more effective methods of helping their students learn will find much of value in the principles and practices of authentic assessment. Teachers will also find that much of what they do already fits comfortably with the strategies suggested in this chapter and elsewhere in this book.

## REFERENCES

Clay, M. (1985). *The Early Detection of Reading Difficulties*, 3rd ed. Portsmouth, N.H.: Heinemann.

Cohen, M. (1980). *First Grade Takes a Test*, illustrated by Lillian Hoban. New York: Dell (Greenwillow).

Crafton, L.K. (1991). *Whole Language: Getting Started . . . Moving Forward*. Katonah, N.Y.: Richard C. Owen Publishers.

Graves, D.H. (1983). *Writing: Teachers and Children at Work*. Portsmouth, N.H.: Heinemann.

Harp, B. (1991). *Assessment and Evaluation in Whole Language Programs*. Norwood, Mass.: Christopher-Gordon Publishers.

Johnston, P. (1991). *Constructive Evaluation of Literate Activity*. New York: Longman.

Kemp, M. (1989). *Watching Children Read and Write: Observational Record for Children with Special Needs*. Portsmouth, N.H.: Heinemann.

King, D. (1993). "Assessment and Evaluation in Bilingual and Multicultural Classrooms." In *Assessment and Evaluation in Whole Language Classrooms*, edited by B. Harp. Norwood, Mass.: Christopher-Gordon Publishers.

Neill, D., and N. Medina. (1989). "Standardized Testing: Harmful to Educational Health." *Phi Delta Kappan* 70, 9: 688–697.

Paulson, F., P. Paulson, and C. Meyer. (February 1991). "What Makes a Portfolio a Portfolio." *Educational Leadership* 48, 5: 60–64.

Rhodes, L., and S. Nathenson-Mejia. (1992). "Anecdotal Records: A Powerful Tool for Ongoing Literacy Assessment." *The Reading Teacher* 45, 7: 502–509.

Routman, R. (1988). *Transitions: From Literature to Literacy*. Portsmouth, N.H.: Heinemann.

Routman, R. (1991). *Transitions: Invitations: Changing as Teachers and Learners K–12*. Portsmouth, N.H.: Heinemann.

Sharp, Q. (1989). *Evaluation: Whole Language Checklists for Evaluating Your Children*. New York: Scholastic.

Siu-Runyan, Y. (1991). "Holistic Assessment in Intermediate Classes: Techniques for Informing our Teaching." In *Assessment and Evaluation in Whole Language Programs*, edited by B. Harp. Norwood, Mass.: Christopher-Gordon Publishers.

Shepard, L. (1989). "Why We Need Better Assessments." *Educational Leadership* 46, 7: 4–9.

Tierney, R.J., M.A. Carter, and L.E. Desai. (1991). *Portfolio Assessment in the Reading-Writing Classroom*. Norwood, Mass.: Christopher-Gordon Publishers.

Wiggins, G. (1989). "A True Test: Toward Authentic and Equitable Forms of Assessment." *Phi Delta Kappan* 70, 9: 703–713.

Winograd, P., S. Paris, and C. Bridge. (1991) "Improving the Assessment of Literacy." *The Reading Teacher* 45, 2: 108–116.

# I-9

# Three Pictures of an Exhibition: Warm, Cool, and Hard

JOSEPH P. MCDONALD
COALITION FOR ESSENTIAL SCHOOLS

First published in *Phi Delta Kappan* 74, 6 (February 1993): 480–486. Reprinted with permission from Phi Delta Kappa and the author.

# Three Pictures of an Exhibition: Warm, Cool, and Hard

*Mr. McDonald takes a close look at alternative assessment as practiced in a large-city high school and draws some instructive tips from what he sees.*

BY JOSEPH P. MCDONALD

Illustration by Mario Noche

MARY McLEOD Bethune High School,[1] situated in a large American city, is a partner in the high school reform effort that was launched by the Coalition of Essential Schools in 1984. The Coalition advocates simple goals that apply to all students and that are focused on intellectual achievement. It also advocates a pedagogy that combines a personalized and caring environment with a focus on student production – on the "student as worker." It argues for "graduation by exhibition" rather than graduation as the consequence of time served in class. This phrase – "graduation by exhibition" – leads a double life in Coalition discourse. It refers to a method of assessment: one that measures valued knowledge directly rather than indirectly; that demands performance, the application of knowledge, and the use of metacognitive strate-

---

*JOSEPH P. McDONALD is a senior researcher with the Coalition of Essential Schools, Brown University, Providence, R.I.* *This article is based on a paper presented at the 1991 annual meeting of the American Educational Research Association, held in Chicago. The research it reports was funded by the IBM Corporation and the UPS Foundation.*

gies; that happens in some public setting; and that guards a gate to graduation. At the same time, the phrase also refers to the Coalition's own version of outcomes-based education, an ideal of schoolkeeping in which all systems are oriented toward producing graduates in the image of the school's collective vision of competent intellectual performance.

The faculty at Bethune High School seeks to produce graduates who demonstrate imagination, power, confidence, and compassion in intellectual matters; who communicate skillfully; who have not just textbook knowledge of the major subjects they have studied but also can apply "content" to investigations of the important issues of their time. "Planning backward" from this compelling image, the school's faculty members have taught themselves to teach in interdisciplinary teams, and they have given their students responsibility, gotten to know them well, asked them to find and solve problems, taught them how to present themselves, trusted them to work on their own, offered them time and freedom to work outside the walls of the school, and believed in them.

I visited Bethune on its first Exhibition Day, a month before graduation. I was there to witness alternative assessment in practice. I was also there to see this high school's standards in operation and, along with the rest of the school, to test for a match between its graduating seniors and the image, crafted four years earlier, of what they might be. My visit gave me three pictures that I carry with me when I evaluate other schools that follow Bethune's lead – pictures I characterize as warm, cool, and hard. Each picture conveys a key dimension of Bethune's achievements as well as some important lessons for followers.

## WARM

Meeting some seniors on the morning of the school's first Exhibition Day, I can feel the energy they feel. Though Bethune High School is new to me, I am certain from its air alone that it is different today – crazier, but more focused too, like backstage just before the curtain rises. Perhaps the teachers are also more forgiving of interruptions: one exhibitor's search for a missing poster, another's panicky recruitment of substitute models for her fashion show, even my late arrival.

The first of the senior exhibitions will begin in 20 minutes, culminating a process that began formally a year ago. Each senior has defined an interdisciplinary question and pursued it through library research and field work. Each has met the deadlines for a series of submissions: a list of research sources, an outline, note cards, an introduction, a whole draft, a finished paper. This is routine stuff for high school students – though not usually required of every one of them. But now they are facing something beyond routine: each senior has been allotted a half hour in one of the series of exhibition sessions starting today. During this time, the senior presents highlights from his or her inquiry and then stands for questions. The questioners include teachers, other seniors, underclass students (all of whom attend at least one session of these "orals"), some parents, and some outside visitors like me.

I shake hands all around, appearing confident that what I am about to witness will be interesting, novel, worth my train ride. The confidence is genuine but also a gesture of connection to the proud and friendly company. Privately I am also skeptical, the victim of a number of shoddy school drama productions, art exhibits, and science fairs. These efforts necessarily become invested by a school's belief in itself. That belief is what carries everyone through the burdens of mounting such efforts outside the ordinary bounds of business, though it sometimes swamps critical judgment.

Now for the test. Can each one of the real students live up to the collective image? This is the assessment question. It presents a challenge interwoven with ceremony, rather like an intellectual senior prom. The challenge is real and difficult, which is why these seniors are rushing about, why they are wearing dresses and ties, why they have taken such care with their graphic displays. Still, they have been prepared well for the challenge; they have been rehearsed and coached. That is why parents, teachers, and the principal sit confidently, and why presenters step to the podium and speak confidently. "Good morning," each begins.

It is a scene primed for warm attachments – those investments of belief in the performer that arise from a caring history. Without the benefit of warm attachments, most students do not perform at peak ability. They will prefer, instead, to settle for less than their peak performance and so avoid the risk that their peak may disappoint themselves and others. Thus does assessment of any kind become inauthentic.[2]

> **I shake hands all around, appearing confident that what I am about to witness will be interesting, novel, worth my train ride.**

Of course, warm attachments can also kill authenticity by suppressing participants' critical perception. So warm attachments must be balanced by cool judgments. For example, teachers must balance their warm sense of how far each student has come with a cool sense of what the outside expects of the young educated adult. This is tough work in any setting – particularly a ceremonial one.

Still, a number of teachers try to maintain some objectivity on this first day of the Bethune exhibitions. They ask hard questions. I do, too, though I find myself so pleased by the presenters' fortitude in standing for such questions that I am perhaps too appreciative of the content of the answers. I ask the senior who has just spoken about African-American achievements in several fields but has omitted the arts to correct the omission.

"Who are the African-American writers you admire?" I demand. "Maya Angelou, for one," he answers immediately and goes on to say why in detail, though for my part he might well have stopped at the name alone, so impressed am I with his presence of mind. And I am so taken by the clarity and potential of another senior's exhibition question – Is our water safe to drink? – so warmly invested in my own sense of its generative power and its potential interdisciplinary sweep, that I fail to recognize that it is only my own question that moves the exhibitor's focus past chemistry and public health to an observation about water and images of purity.

Meanwhile, amid applause and the palpable pride of teachers, students, and parents, another test is under way: What do these performances reveal about the school's own systems? This is the standard-setting question. And, like the assessment question, it demands both warm and cool approaches.

I wonder, as I sit observing, whether the school will dare to face this question. When emotions have subsided, for example, will the faculty review with a critical eye the videos it has made of these exhibitions? Video is, as they say, a cool medium. Will younger students watch them too? Will the portfolios that must accompany the exhibitions (documenting phases of the inquiry) inform teachers' and students' planning of next year's courses and projects? How far back will the influence of these exhibitions and portfolios run? Will the experience of next year's ninth-graders be different because of what this year's seniors have done? Will there be more time set aside from now on for project work and coaching groups? Will there be more practice speaking before strangers? Will the school be materially affected – that is, will structures and resources be shifted?

## COOL

A year later I watch a video of the previous May's exhibitions at Bethune High School. The video contains a selection of four exhibitions, none of them from the session I attended. The school made the selection for its own purposes – to explain itself to interested outsiders, to prepare students to follow the first Exhibition Day's pioneers, and to exercise its critical eye.

I watch the tape from my perspective as a researcher, interested to see by its light what I take to be some key dimensions of assessment by exhibition. Of course, sitting in my living room, distant in space and time from the warm attachments I felt when I was present at the session, I am primed for cool judgment.

The four exhibitions on this tape all deal with contemporary topics: substance abuse and fetal development, child abuse and the juvenile court system, recycling and the environment, cholesterol and the American diet. These are my characterizations of the topics, and I phrase them as I do – simple intersections of relatively broad subjects – both to emphasize the immensity of the task the exhibitors faced and to say that each of the exhibitions might have benefited from more focus. This is the way with contemporary topics. They are difficult to delineate, which makes them unlike the topics drawn from textbooks that high school students more commonly face. And even when one has delineated a contemporary topic reasonably well, that topic may nonetheless slide away even as one speaks.

This happened to the exhibitor who addressed child abuse and the juvenile court system, but who seemed through most of his exhibition to focus solely on the court system. Only the first questions from his teacher, intent on directing his attention to child abuse, cued me that that had indeed been an emphasis of his rather nervous, muddled introduction. I think I must have lost track of this emphasis when I sensed the confidence that emerged as he turned, pointer in hand, to his first exhibits: a flow chart description of the juvenile court system and a collection of documents from key junctures in that system. He himself may have lost track there also.

In fact, I noticed as I watched the other exhibitions that a strong presence sometimes overwhelmed substance. That may be a chronic problem in exhibitions generally: they require courage, and courage – as rare in school as elsewhere – dazzles. So one exhibitor's entertaining demonstration of how to cook low-fat crab cakes, worthy of midmorning network television, seemed to distract the exhibitor herself from attending to her main concern, the impact of cholesterol-laden foods on health. I think it distracted the audience, too, which got to eat the crab cakes. Another exhibitor's cleverness in presenting his own videotape of Earth Day events seemed to relieve him in his own eyes – and perhaps in those of the audience as well – from having to tie this exhibit to his main point. These lapses would have been far less apparent to me, of course, from a warm point of view, had I been in the audience anticipating crab cakes, impressed by an exhibitor's use of visual media.

What can a school do to guard against such lapses? The answer, I think, is, "Nothing much within the warm setting." Oral exhibitions like those at Bethune High School feature substance *and* presence – but, I suspect, they typically tip the balance in favor of the latter. For one thing, contemporary American students, enveloped by pop culture, have enormously greater exposure to models of presence than they have to models of substance. But, taking advantage of cool opportunities, the school can and must ensure that the assessment system, of which the public exhibition is merely one important part, builds in plenty of substance checks along the way, as well as plenty of models.

Bethune's requirement of a portfolio documenting the inquiry – including note cards, interview transcripts, drafts, a finished paper, and so on – is an example of a substance check. The trick in using portfolios is to make them "count" as much as the presentation does, to design them so that they rigorously document the research process. Any teacher who has ever collected a novice researcher's note cards knows how easily such a process can become an exercise in assessing form rather than substance. Without taking the time to interview the student, it is often impossible to know what the cards really represent of the process under way. To be worth the trouble it takes to keep and review it, a portfolio of this kind will be more than the documentation of a performance; it will be the documentation of a *coached* performance – an artifact of teaching as well as learning.[3]

Coaching, however, involves more than ongoing assessment and support in following some given process. It also requires the provision of models – the elaboration by various means of a vision that the student may aim for. This must be more than a procedural vision – a model of how to find and follow *any* compelling question. Seventeen-year-olds must also dig for substance in the gardens of mentors who can show them in various ways what excellence means.[4] Teachers cannot know enough about all the fields that seniors may take an interest in. Therefore, they cannot be the sole suppliers of question-generating experience, the sole models of inquiry, or the sole guardians of standards.

Of course, teachers can and must model inquiry in general, but good high school inquiry has more than generic quality. It has the whiff, as well, of incipient expertise. This is what we are used to sensing in some "extracurricular" pursuits of teenagers – the knowledge of cars or music or perhaps retail merchandising that already exceeds what we may know about these things and that we sense is headed toward real depth. We seldom expect to find the same quality in schoolwork. One reason, I think, is that we tacitly limit schoolwork to school grounds. Bethune and schools like it seem willing to give students access to many gardens on other grounds. The success of the effort will depend on whether the students' digging is well supported – both on site and back at school.

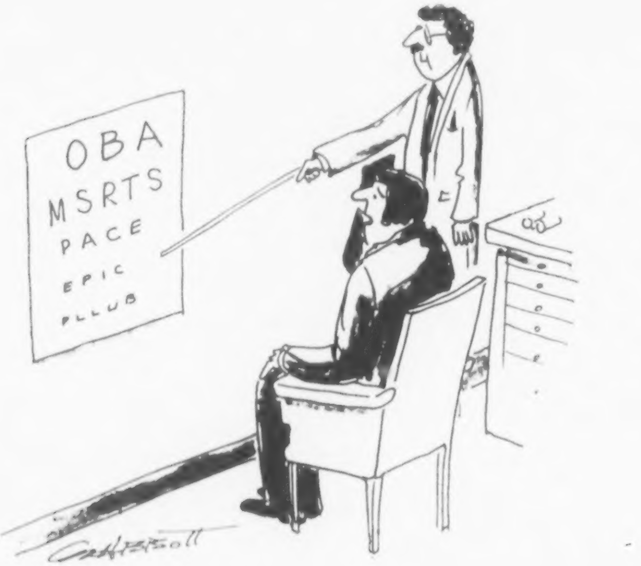
Meanwhile, there are other kinds of gardens in which future exhibitors must dig. These gardens are any relevant and rich sources of data, insight, and artifact. Access can be in person – for example, by means of long and structured visits to museums and libraries or a series of interviews – or it can be by means of technology – books and microfiche or telecommunications and CD-ROM.

But access alone is far from enough. It has become routine to talk about an information explosion and its challenge for school curricula. But to understand the real thing, take any of the four topics addressed in the tape of the Bethune exhibitions – let's say, substance abuse and fetal development – into a well-equipped library. See for yourself what the on-line catalog, the electronic databases, and the periodical indexes provide. Unless you are highly focused in your inquiry and comfortable with all the odd sorts of discourse about data in the fields this topic touches, you will surely feel overwhelmed – you may even yearn for the comforting information vacuum of the average high school library with its out-of-date books and its incomplete collections of *Time* and *Science News*. So the Bethune exhibitors need access to rich information and also to teachers willing to take their coaching cues from information they did not provide and do not control. Moreover, the exhibitors need these things long before they face the pressures of Exhibition Day.

This is an enormous challenge for Bethune or any other high school, but Bethune is ahead of most others in daring to face it. Watching the video and prompted by these exhibitions, the staff can begin to imagine – as I do – necessary shifts in instructional design. Signs of what is needed show up in these exhibitors' occasionally shaky command of information, their apparent gaps of understanding, the unaccountable leaps they sometimes make. "Smoking causes low-weight babies," one exhibitor declares baldly, and "fetal alcohol syndrome leads to mental retardation." ("*Can* lead to mental retardation," I want to add from my cool living room.) "The three causes of retardation . . . ," he begins, and I wish I could interrupt coolly, "Only three?"

I offer these phrases as snapshots of a student trying to hold his head above very deep water. He has a lively intelligence that shows even on the cool video, but he needs more time among adults who swim expertly in this same water. As flexible as Bethune's structures have been, this student needs even more time *outside* the school to do school-sanctioned and school-supported work. But this puts great pressures on his school: if he spends so much time in the field, how can he complete all his course requirements? And if the school is not to discharge its support functions through coursework, then how may it do so?

This student also needs even more access – coached access – to the kinds of data he proudly supplies in his exhibition – city, county, and state data – but that he interprets more clumsily than one would wish. It is not just that he needs more time and opportunity to visit government offices; he probably also needs more coaching on how to read what such foraging turns up. This may also require even more of a shift than Bethune has already made toward coached project time and what might be called "studio work."



*"Outcomes-Based Accreditation, Migrant Student Record Transfer System, Policy Analysis for California Education, Educators Promoting International Cooperation . . ."*

---

**HARD**

A warm regard provides the social investment that both students and schools need to develop, while a periodic cool regard provides the critical distance they need to reflect on this development and to shape it. But warm and cool are not enough. What is also wanted is some direction for development: something hard that exists apart from warm or cool relationships and that orients them – a framework of standards.

In matters of assessment, there are three principal standards: validity, reliability, and equity. Though these terms seem alien to school life, as if they invariably involved psychometric engineering, the concerns that inhere in them are nonetheless vital to school development.[5] As I say, they orient it. The third picture of Bethune's work, framed by these hard standards, tries to capture this orientation.

*Validity.* According to the *Standards for Educational and Psychological Testing* of the American Psychological Association, validity refers to the appropriateness, meaningfulness, and usefulness of inferences based on assessment.[6] I like to think that validity constitutes the interface of assessment and the world: it demands of the test that it measure what is really valued.

Evidence of validity in assessment typically comes in three varieties. The first, "construct-related," has to do with the relationship between the assessment tasks and whatever abstract phenomenon is of interest – ability to write, for example, or to conduct an inquiry or to handle questions on one's feet. The second, "content-related," has to do with the relationship between the full sample of the assessment's demands and the complexity of the domain it purports to represent. And the third, "criterion-related," has to do with the relationship between an assessment's findings and some other measure – for example, between how well students do on one test and how well they do on another or between how well they do on one test and how well they do in college.[7]

Psychometricians weigh these kinds of evidence by means of various logical and empirical procedures involving item analyses, the testimony of experts in the domain, and tests of the test. From the perspective of Bethune High School, the methods seem esoteric at first glance, but the glance is deceiving. In fact, the faculty at Bethune does analyze its exhibition requirements to determine their match with its real values. It does invite the oversight of outsiders (the example of this article indicates that). Finally, it should and does corroborate exhibition results with other indicators. One member of the faculty, for example, has toured the campuses where many of the former exhibitors now study, interviewing them in order to evaluate how well their school study and exhibitions prepared them for college.

From the perspective of Bethune High School, however, a concern for validity in assessment must involve much more than the local adaptation of psychometric techniques. The heart of the matter is the assurance of connection between school and the world. Much of what I say above – both when I write warmly and when I write coolly – has to do with the scope and quality of this connection. Are the students immersed in the real stuff? In the way they write and speak and handle questions, do they show the right stuff? Validity in the assessment by exhibition at Bethune High School cannot be established once and for all; it is a matter of relationship, of a willingness on the part of the school to open its doors to the outside, so that mentors and critical friends may visit and so that students can go out digging. It is a matter, as I suggest above, of faculty courage. Happily, this is plentiful at Bethune, so the picture is heartening.

*Reliability.* Packaged tests usually advertise their reliability in correlation coefficients: measures of how well results from one form of the test match those from another form or how well results from even-numbered items match results from odd-numbered items or how well results from tests match results from retests. From the perspective of Bethune High School, this may seem unimportant – but, again, a quick glance is deceiving. Actually, the long-term success of the assessment system there will depend on reliability in at least two respects: 1) the achievement of interrater agreement (among the multiple judges of the exhibitions and portfolios) and 2) the achievement of consistency in scoring (across the various Exhibition Days, between one year's exhibitions and the next, between one student's exhibition and the next student's).

The main difference, however, between achieving assessment reliability within a school and achieving it in an external test or assessment system is that the former is both a technical *and* a structural problem. The heart of reliability at the school level is the assurance of communication among teachers, and this depends on structure. In fact, no standard-setting devices, interrater agreement mechanisms, or consistency checks can work inside the ordinary cellular structure that is characteristic of nearly every U.S. high school. Without new structures, teachers will remain isolated, and assessment will remain private and unreliable. And restructuring may not be enough, either. A restructuring plan, for example, may address other important needs – perhaps providing collaborative opportunities for planning and teaching – yet still leave assessment private or else bump it up to some "higher" level (a supposed exchange of power for "accountability"). The latter action may well leave schools and students worse off than before. As Vito Perrone notes, accountability, to be worthy of the name, must be rooted directly in teaching, not float above it.[8]

In an important sense, reliability, from the perspective of Bethune High School, is less about seeking agreement – at least initially – than it is about seeking and sustaining productive disagreement. Before they can achieve the degree of reliability that their students and their bold experiment deserve, the teachers at Bethune must dare to disagree with one another and to abandon the safety of their own private grade books. They must invent methods to assess together, as well as teach and plan and advise students together. They have made great strides in this direction, including team-teaching arrangements and the provision of common planning and meeting time. Again, the picture is heartening.

*Equity.* Equity in assessment is often equated with the elimination of bias. Again, this is a matter for psychometric engineering: one determines by various means which test items cause particular categories of test-takers to stumble disproportionately relative to their performance on the test as a whole. So boys have trouble with items that refer to colors, and girls have trouble with items that refer to sports. This method of bias detection works because the test items that fall under suspicion are, after all, indirect indicators of the achievement that is actually of interest. The item that refers to colors or to sports really seeks to measure reading comprehension or the ability to solve mathematical word problems; the content of the item – as contrasted

> **From the perspective of the school, an assessment by exhibition — in all its dimensions — is a systemic thing.**

with its objective – is irrelevant or, in these cases, misleading.

In performance-based assessment, however, one strives for direct measure, which is to say that content counts. The result is that the detection of bias – while nonetheless important – must be merely one step in achieving equity. What does the Bethune faculty do if it discovers, in a careful review of exhibition results, that boys tend to do better in fielding questions than girls? The answer, I think, would not be "eliminate the questions," as it might be if the format did not matter, if the faculty had not deliberately intended that all its graduates would be able to handle such questions successfully. Instead, I think the school's response must be to build a better scaffold for the girls.

In general, equity in assessment at Bethune High School is a matter of scaffolding. The central problem is how to get everybody to a place where some never imagined that they could get. This is equity of results rather than merely of processes, but, because what is wanted is an equity of optimal rather than minimal results, new and exceptional processes are required. To continue the example begun above, some girls will surely not get there if they are treated just like boys at every step of the way. Nor will they get there if, for even a moment, they are less pushed than the boys or if, by various means, they are segregated from the boys. What is wanted is a community of striving in which the standards of achievement are plainly visible and plainly applicable to all, but in which support, time, and structure all vary according to need.

Of course, such a community is much easier to describe in abstract terms than it is to build. The fact is, at this early point in the effort to remake the American high school, we still lack good "architectural" models. Nevertheless, we know some of the building materials we will need. For example, we will still need "courses." But, as I suggested above, we will need something else besides – something like "projects," something like what the Bethune students do in their senior year. And we'll need curriculum. But we'll need a kind of curriculum that is not wholly contained inside "subjects," one that links rather than isolates learning experiences, one more embedded than we can now imagine in real contexts. We will need mechanisms that enable teachers to tune their judgments to common and high standards.[9] These may take the form of interrater reliability schemes (both within the school and between the school and the outside), which function to aim instruction high as well as to keep assessment reliable.

Then we will need new systems of information management that are far more responsive and capacious than school libraries, transcripts, and textbooks. We'll need grouping patterns more respectful of human diversity and potential than age-grading and tracking, and we'll need management systems that can sustain community better than hierarchy can. Finally, and most important, we will need a commitment to democratic principles that is more than platitudinous – that can materially affect the school and pervade its habits. At the very least, this will involve a new conception of intelligence[10] and an uncommon tolerance for difference and conflict.

If the preceding paragraph has seemed to veer far away from the issues of assessment and also from Bethune, the effect is intended. From the perspective of the school, an assessment by exhibition – in all its dimensions – is a systemic thing. It cannot be a matter of grafting something new and clever onto an old design – a new portfolio system that is supposed to make everything different, though it has been dumped onto the school's overworked and isolated English teachers, or a new "authentic" state or national exam that is supposed to force new modes of instruction but that, as Lorrie Shepard points out, also implicitly cheats teachers of their professional status.[11] From the perspective of Bethune High School specifically, the prospect of an utterly redesigned high school is a distant and daunting one, yet it is exactly what attracted the school to exhibitions in the first place. And this is the most heartening picture of all: that Bethune and some other daring schools would be willing to scout for the rest of us.

---

1. The name of the school is a pseudonym, meant to protect the privacy of the exhibitors.

2. I take *authenticity* in assessment to be less a matter of method and more a matter of impact: does the assessment assess what it means to, and is the one assessed genuinely engaged?

3. Lee Shulman describes a portfolio as the documentation of a coached performance and adds that the coaching does not compromise the use of the portfolio as an assessment device. Does the dissertation advisor, he argues, claim to have had no influence on the advisee for fear of compromising the judging of the dissertation? I would add that the juxtaposition of coaching and judging, while natural and appropriate, also demands structural arrangements that share the intellectual and emotional load. Thus we have dissertation committees in the university, and we need something like them in the high school.

4. Howard Gardner, *To Open Minds: Chinese Clues to the Dilemma of Contemporary Education* (New York: Basic Books, 1989).

5. Richard J. Stiggins, "Assessment Literacy," *Phi Delta Kappan*, March 1991, pp. 534-39.

6. *Standards for Educational and Psychological Testing* (Washington, D.C.: American Psychological Association, 1985).

7. Other kinds of validity than those mentioned have recently been proposed. Among the most relevant of these to the issues addressed here is *systemic validity*, which has to do with the corruptibility of the assessment: if teachers teach to this test, will students be better or worse off? (See John R. Fredericksen and Allan Collins, "A Systems Approach to Educational Testing," *Educational Researcher*, December 1989, pp. 27-32.) This is related to what Lauren Resnick and others have called *consequential validity*, which has to do with the uses to which a test is put.

8. Vito Perrone, "Testing, Teachers, and Schools," in Judah L. Schwartz and Katherine A. Viator, eds., *The Process of Secrecy: The Social, Intellectual, and Psychological Costs of Current Assessment Practice* (Cambridge, Mass.: Educational Technology Center, Graduate School of Education, Harvard University, 1990), pp. 99-107.

9. Grant Wiggins, "Standards, Not Standardization: Evoking Quality Student Work," *Educational Leadership*, February 1991, pp. 18-25.

10. Dennie Wolf et al., "To Use Their Minds Well: Investigating New Forms of Student Assessment," in Gerald Grant, ed., *Review of Research in Education*, vol. 17 (Washington, D.C.: American Educational Research Association, 1991), pp. 31-74.

11. Michael W. Kirst, "Interview on Assessment Issues with Lorrie Shepard," *Educational Researcher*, March 1991, pp. 21-23.

BEST COPY AVAILABLE

# I-10

# Keeping Performance Assessment in Perspective

Richard J. Stiggins
Assessment Training Institute

## Introduction

Performance assessment is one of several valuable tools for gathering information about student achievement—but we must keep it in perspective by remaining aware of the array of educational outcomes to be assessed and of the keys to assessing each. An analysis of the match between outcomes and assessment methods reveals both the strengths and limitations of performance assessment.

In recent years, some have contended that performance assessments replace objectively scored paper-and-pencil tests because of the rich detail they provide about student achievement. This contention is based on the assumptions that these two methods are interchangeable in the outcomes they reflect, and performance assessment is better at assessing those outcomes. In this paper, I argue that these assumptions are invalid. Performance assessment and objectively scored paper-and-pencil tests attend to different outcomes. One is not superior to the other: They are equally powerful in fulfilling their respective assessment roles—in the hands of qualified users.

I also argue that objectively scored paper-and-pencil tests and performance assessment are not the only options available. We have such options as essay assessments and direct personal communication with students, and we need to take advantage of all options to assess relevant outcomes effectively. All assessment alternatives can help us reach different assessment goals.

## The Full Array of Valued Outcomes

In most classrooms, five interrelated outcomes are important. We expect students to:

- Master substantive subject matter knowledge
- Use that knowledge to reason and solve problems
- Demonstrate achievement-related skills, such as read aloud, speak in a second language, and perform psychomotor acts.
- Create achievement-related products, such as samples of writing and art products.
- Attain feelings or affective states, such as attitudes, interests, and motivational predisposition.

These categories are useful in thinking about assessment because each presents different assessment challenges. But first, let's understand the categories of outcomes more thoroughly.

---

From *Student Centered Classroom Assessment*, by Richard J. Stiggins. Copyright 1994 by Macmillan Publishing Company. Reprinted with permission.

### Knowledge Outcomes

Most teachers expect their students to master relevant knowledge. Typically, this includes facts, relationships, algorithms, procedures, and the like. In fact, all academic achievement is based on substantive knowledge, and none of the four outcomes discussed below can be attained without it.

For example, in every reasoning and problem-solving context, success depends on the student's mastery of the facts, information, procedures, and conceptual relationships required to solve the problem. We cannot solve math problems without knowledge of math facts, number systems, and problem solving procedures. We cannot speak a foreign language without mastery of its vocabulary, syntax, and structure. We cannot write an essay in English without a practical knowledge of letters, words, sentences, paragraphs, grammar, and knowledge of something to write about.

All academic competencies rest on a foundation of knowledge. We cannot read with comprehension if we lack sufficient background knowledge about the material. We cannot write an essay on the Civil War without access to facts about that event. Therefore, at any point in the instructional process, a teacher concerned about student attainment of the building blocks to ultimate reasoning, skill, or product competence might wish to assess student mastery of relevant knowledge.

I did not say the teacher might expect students to "memorize" knowledge (i.e., facts and information). I said master important knowledge—meaning to gain control over the specified material. That control can be gained either through memorization or through the appropriate reference materials.

The world does not function on memorized facts and information. As adults, we quickly learn that we cannot remember everything we need to know. Because we generate new knowledge in almost all fields of study at incredible rates, keeping up with it is impossible.

As a result, we have learned to use computers and other reference technologies to help us gain access to the most current knowledge when we need it. We maintain access to the important material only for as long as we need it—then we let the knowledge go. This is a practical and relevant way to master substantive subject matter.

In short, this knowledge category of educational outcomes includes achievement targets (core facts, information, and relationships)that students memorize, and elements and structures of knowledge that students research as needed. Both are important and both present special assessment challenges.

### Reasoning Outcomes

Teachers rarley want students to master content just for the sake of knowing it. We want students to use their knowledge to reason and to solve certain kinds of problems. For example, we want them to analyze and solve story problems in math, compare political events or leaders of the past, reason inductively and deductively in science, and evaluate opposing positions on social and scientific issues. We want them to use what they know within the problem context to achieve a desired solution.

If we value such outcomes for our students, our first challenge is to precisely define *reasoning*. Some teachers create their own definitions. Others rely on one or more of the many conceptual frameworks scholars have generated over the centuries to define thinking and problem solving processes (e.g., Bloom 1956, Quellmalz 1987, Marzano 1992).

Our second challenge is knowing how to translate our vision into sound assessments. Once we understand how to make those translations, we can use that wisdom either to devise our own assessment or to verify the quality and appropriateness of others' assessments. However, until we possess a vision of the proficient problem solver, we cannot assess that proficiency.

### Skill Outcomes

In most classrooms, teachers want their students to demonstrate certain kinds of skills. For example, at the primary grade level, a teacher might expect certain kinds of social interaction behaviors or might look for the earliest oral reading skills. At the elementary level, a teacher might observe student performance in cooperative group activities. In middle school, manipulation of science lab apparatus might be important. And at the high school level, public speaking or the ability to converse in a second language might represent valued outcomes.

In all cases, success lies in doing well. The assessment challenge lies in defining in words, through example, or both, what it means to do it well.

Again, note that skillful performance requires prior mastery of procedural knowledge about how to perform and the development of the reasoning power needed to use that knowledge across performance contexts. However, in this category, the objective is for the student to demonstrate the ability to put it all together, which is precisely why achievement-related skills often represent complex targets requiring more sophisticated assessments.

### Product Outcomes

Another way to succeed in classrooms is through certain achievement-related products—tangible entities that reveal levels of student achievement through some medium. For example, a high school social studies teacher might have students prepare a term paper. A middle school shop teacher might have students build products. An elementary school teacher might challenge students to prepare their first science lab reports. Or a primary grade teacher might collect samples of student art work. Another classic example of this kind of target that crosses grade levels is the creation of writing samples—tangible products that contain evidence of proficiency.

In all such cases, success lies in creating products that possess certain key attributes—that meet certain criteria. The assessment challenge we face is to define in clear, written descriptions or through example what those attributes are. We must be able to specify exactly how high- and low-quality products differ.

Note once again that successful performance arises out of student mastery of prerequisite knowledge and through the application of appropriate reasoning and problem solving strategies. In addition, in this case, performers have to carry out certain steps in the process of creating the desired product. But evidence of ultimate success lies in the product itself and the product is what we must assess.

### Affective Outcomes

This final category of valued outcomes includes characteristics that go into the realms of student feelings. Many teachers hope that students develop positive self-concepts, positive attitudes toward school, strong values for learning, strong interests, and strong internal control over their well-being. Each of these contains three essential elements: an object or a focus, a direction (positive or negative), and a level of intensity (strong or weak). These are what we must assess.

## The Full Range of Assessment Options

As we plan to assess these various outcomes, we must select from the four different assessment methods discussed in the introduction:

- Selected response, objectively scored tests
- Essay assessments
- Performance assessments
- Assessments that rely on direct communication with the student.

Each provides its own evidence of student proficiency.

### Selected Response

This category represents the classic paper-and-pencil test. Respondents are asked a series of questions, each accompanied by a range of alternative responses from which to select either the correct answer or the best answer. The index of achievement here is the number of questions answered correctly.

Format options within this category include:

- Multiple-choice exercises
- True-false items
- Matching exercises
- Short answer fill-in items ( a brief answer counted as right or wrong).

### Essay Assessments

In this case, the respondents have an exercise or set of exercises requiring the preparation of an extended written answer. They might be asked to answer a question or to provide an explanation of the solution to a complex problem. Open-ended math problems where students must show all work are in this category. The examiner reads the responses and makes subjective judgments about the quality of the responses.

Evidence of achievement is seen in the conceptual content of the examinee's ideas and the manner in which those ideas are tied together. The index of achievement is the number of points attained out of a total number possible.

### Performance Assessments

These are assessments in which the respondent performs a specified activity or creates a product under the eye of an evaluator who observes performance and makes judgments about the level of achievement. Sometimes evidence of achievement is seen in the respondents ability to carry out the proper sequence of activities or to perform in the appropriate manner. The doing is what that counts. The index of achievement is a performance rating or profile of ratings that reflect on levels of quality in the performance. Examples include evaluations of musical performance, assessments of speaking proficiency, assessments of athletic performance, and road tests of driving skill.

In product evaluations, the respondent creates a complex, achievement-related product intended to have certain attributes, and the assessor examines the product for those attributes. The characteristics of the creation are what count, not the process of creating.

Again,the index of achievement is the rating of product quality. Art or craft products are in this category as are exhibits created for science fairs. Another classic example is the writing assessment in which students produce a sample of their writing that is evaluated for proficiency in the use of written language as well as for content.

### Personal Communication as Assessment

One of the most common ways teachers gather information about student achievement is to talk with them. We don't think of this as assessment in the same sense as a multiple-choice test or a performance assessment, but certain forms of personal communication do represent assessments of student achievement.

These forms of personal communication include questions asked and answered during instruction, interviews, conversations, listening during class discussions, conversations with others (like parents or other teachers) about student achievement, and oral examinations. Personal communication certainly does represent a prominent means of assessment in the classroom.

## Matching Targets to Methods

The quality of any assessment is a function of the developer's ability to blend the five basic kinds of achievement targets with the four kinds of assessment methods. As it turns out, the recipes for creating appropriate blends are straightforward.

All of the possible combinations are depicted in Figure 1. Within each cell the nature and viability of the outcome/method match is defined. Please keep three key points in mind:

First, remember that the first four kinds of achievement targets each build on those that precede

FIGURE 1
ALIGNING ACHIEVEMENT TARGETS AND ASSESSMENT METHODS

| **Achievement Target** | **Selected Response** | **Essay** | **Performance Assessment** | **Personal Communication** |
|---|---|---|---|---|
| Mastery of Knowledge | All formats can be used to assess content knowledge | Can serve to assess student mastery of complex structures of knowledge | Can be used to assess effective use of reference materials | Can assess small domains of knowledge when short-term record keeping is required |
| Reasoning | An excellent way to assess some key kinds of reasoning— but not all kinds | Written descriptions of problem solutions can provide a window into reasoning | Can watch student in the process of problem solving and draw inferences regarding proficiency | Ask student to "think out loud" to examine problem-solving proficiency |
| Skills | Can test for mastery of prerequisite procedural knowledge— but not ability | Can ask student to describe complex "how to" procedural knowledge but not ability to do it | Can observe and evaluate skills as demonstrated | A strong match when the skill in question is oral communication proficiency; can ask student to describe and discuss complex "how to" procedural knowledge |
| Products | Can test for prerequisite knowledge of the attributes of quality products | Can tap ability to describe the attributes of a quality product | Can assess<br>(a) proficiency in carrying out the steps required to create a quality product, and<br>(b) the product itself | Can probe procedural knowledge and/or knowledge of attributes of a quality product |
| Affective | Can develop highly structured questionnaire items | Can use open-ended questionnaire items | Can infer affect from observation of behaviors or examination of products | Can talk with students about their feelings |

BEST COPY AVAILABLE

it. Knowledge is the foundation. Problem solving involves application of knowledge. Skills represent knowledge and problem solving at work. And, quality products are created by using knowledge, thinking, and skills in the right ways. Teachers who are concerned with building competence might need to assess any of these elements of success along the way to ultimate competence.

Second, remember that assessments provide external, visible, and evaluatable indicators of the inner mental state of the respondent. We use assessments to gather evidence of proficiency from which we draw inferences about whether or not the target has been mastered. One key in selecting appropriate assessment methods is to choose one that provides the most accurate approximation of our intended outcome.

And third, remember that a sound assessment relies on a sample of exercises systematically representative of all those possibilities. Based on student performance on the sample, we can generalize how they would have done on an infinite array of possibilities.

Our goal in assessment design is to use the most powerful assessment we can afford. We want the highest resolution picture of real achievement with the smallest possible sample of exercises to give us maximum information at minimum cost.

As you review Figure 1, I hope you come to understand why it is so crucial to understand the achievement target before selecting an assessment method. Different targets require—they don't just permit—the application of different methods.

Every cell contains a potentially useful assessment option but that does not mean all methods can be used to assess all outcomes. Without doubt, some cells depict stronger matches than others. But all methods can contribut some useful information about some of the building blocks of the ultimate outcome. While this may not be important as we design summative assessments of final outcomes, teachers conducting formative assessments of growing students will find it critical.

Given all these possible matches, how do we decide which method to use in any particular instance? In the best of all possible worlds, where both time and assessment resources are unlimited, choose the column that affords the most accurate depictions of the desired outcome. But, because none of us lives in such an ideal place, we must select the methods of assessment that come closest to representing our valued outcome and at the same time fit the resources of our classrooms. As it turns out, we can work some pretty good trade offs between fidelity and cost of assessment if we know what we're doing. To illustrate, move across the table left to right, covering one row at a time.

## Assessing Knowledge

Obviously, we can use the selected response objective paper-and-pencil test to measure student memorization of facts and information. Typically, these assessments tend to test knowledge of elemental facts in isolation, such as facts about the Civil War, spelling, vocabulary, and earth science. These tests fit efficiently into the realities of most classrooms.

However, when the domain of knowledge is defined as larger structures of knowledge where elements are related in complex ways, we can test student mastery using the essay format. Examples of what students might be asked to know include the causes of the Civil War, or differences among indigenous, metamorphic and sedimentary rocks.

With respect to the use of performance assessment as an index of student mastery of content knowledge, remember that we must all become competent users of resource materials as one means of gaining control over useful knowledge.

For this reason, I have expanded my own thinking about mastery of content knowledge to include the ability to use reference materials to retrieve needed information. One can use performance assessment methodology—observing students in the process of using references resources—to determine their proficiency in this performance arena.

However, using a product evaluation as the basis for assessing mastery of content knowledge presents a special kind of problem having to do with the nature of the inferences based on student performance. Where the student creates a high-quality product, sample evidence exists that the performer mastered the necessary foundational knowledge.

But difficulties begin to arise when we are confronted with a poor-quality product. Standing by itself, that product does not reveal what went wrong. Perhaps the student had not mastered the foundation knowledge. Or perhaps something in the process of creating the product went haywire. Or perhaps the student possessed the requisite knowledge and understood the process, but didn't try. So we see rival explanations for the outcome. If we infer that the problem arose from a lack of knowledge, we might well be wrong. Without following up with other assessment methods,we just don't know. And we need to know what went wrong so we can support the student's attempts to improve.

If follow-up assessment using selected response, essay, or personal communication-based assessment methods reveals a lack of knowledge, then we know what to do. On the other hand, if it reveals a lack of motivation, then a completely different course of action is called for. Without the follow-up, we just don't know what to do. Therefore, I recommend against stand-alone performance assessment as the basis for the assessment of student mastery of knowledge outcomes.

The final option for assessing mastery of knowledge is direct personal communication with the student—for example, asking questions and evaluating answers. This is a good match—especially with limited amounts of knowledge to be mastered, few students to be assessed, and in contexts where records of performance need not be stored for long periods of time.

I impose these conditions because this is a labor-intensive assessment method. So, if the domain of knowledge to be assessed is large, the user must ask a large sample of questions and that doesn't fit the resource realities in most classrooms. Further, if many students will be assessed, this option may take too much time to sample each student's achievement in a representative manner. And, if records of performance must be stored over an extended period of time, written records will be needed for each student over a broad sample of questions. This takes a great deal of time and energy.

Assessment via personal communication works best when the teacher makes a quick check during instruction of student mastery of a few critical facts in order to make any needed adjustments.

### Assessing Reasoning

In most classrooms, teachers want students to use the knowledge they master in productive ways to reason and solve complex problems. We can define what it means to be a proficient reasoner and problem solver in many ways.

We value critical thinking—the ability to make an evaluative judgment and defend it through the rigorous application of standards or criteria. Just as movie or restaurant critics evaluate according to criteria, so too can students evaluate the quality of a piece of literature or the strength of a scientific argument.

We value inferential thinking—the ability to draw generalizations from evidence or to predict the outcomes of some set of circumstances. And yet another is to use foundational knowledge to compare and contrast things—to infer similarities and differences.

How does one assess these kinds of outcomes? Our four methodological choices all provide excellent options for teachers who possess a clear vision of the kinds of thinking they wish to assess and sufficient craft knowledge of those assessment methods. For example, we can use selected response exercises to reach certain kinds of reasoning—but not all. We can use them to see if students can analyze things, compare them, or draw inferences; to evaluate math problem-solving proficiency; or to assess procedural knowledge of how to use technology in science.

Selected response exercises are more flexible than most people think. But with limits. Evaluative thinking—expressing and defending a point of view—cannot be tested using multiple-choice or true-false items because it requires at least a written expression of the defense.

Further, when the problem is multifaceted and complex, involving several steps, the application of several different kinds of thinking and several problem solvers working together—as real-world problems often do—then more complex assessment methods are needed.

Essay tests are an excellent way to assess more complex student reasoning and problem solving. Student writing provides a window into student reasoning. In fact, students often can be encouraged to look through this window to become assessors of their own reasoning proficiency. Teachers can devise challenging exercises that ask students to analyze, compare, draw complex inferences, evaluate, or use some combination of these.

Of course, the key to evaluating the quality of student responses to such exercises is for the assessor to understand the kind of thinking required and to know what to look for. A clear and appropriate set of scoring criteria is essential.

Performance assessment represents yet another excellent option in this performance arena. We can watch students in the act of problem solving in a science lab, for example, and draw inferences about their proficiency. To the extent that they carry out proper procedures, they reveal their inner reasoning processes.

However, drawing inferences about reasoning proficiency based on the quality of student products can be risky. Certainly, a high-quality product evidences sound thinking. But in the classroom, where assessment is used to promote student development, we run into difficulty explaining and helping with weak performance. Did the student lack foundational knowledge, fail to think and problem solve productively, or lack motivation? Without follow-up assessment by other means, we just don't know. If we don't follow up with supplemental assessment our remedy will be inefficient. We may waste valuable time reteaching material already mastered or teaching thinking skills already developed.

And finally, one of the strongest matches between a target and an assessment method on the entire table is the use of personal communication to evaluate student reasoning. Teachers can:

- Ask questions that probe thinking
- Have students ask each other questions and listen to both the questions and the responses
- Have students describe aloud what they are thinking as they confront a problem
- Have students recount their thinking processes

- Listen during class discussions for evidence of sound, appropriate thinking.

Just talking informally with students can reveal so much, when we know what we're looking for.

### Assessing Skills

When our assessment goal is to see if students can demonstrate certain kinds of achievement-related skills, such as speaking in a second language, giving a formal speech, or interacting with classmates in socially acceptable ways, we have just one option: observe them while they exhibit the desired behaviors and make judgments as to their quality—performance assessment. We have no other choice. Each of the other three assessment options falls short.

But sometimes limited resources make it impossible to assess the real thing. At those times, we may need to go for second best. Here is what you get when you trade some of your fidelity for greater efficiency in the assessment of skills:

You can use selected response test items to learn if the student can recognize the right behaviors. For example, given a number of demonstrations of performance (on video, perhaps) can the respondent identify the best performance? Or we can use a multiple-choice item format to see if students know the proper sequence of activities when that's a relevant facet of the outcome. Given several descriptions of a particular procedure, can the respondent identify the proper method? Or we can use this method to ask if students have mastered the vocabulary needed to communicate about the desired skills.

Just realize, however, that these represent important prerequisites to effective performance—the building blocks to competence. But such items cannot assess that examinee's actual level of skill in performing.

In this same sense, students could write an essay about the criteria they might use to evaluate performance in a vocal music competition, and that knowledge might well represent an important foundation for being able to perform well in such a competition. But this will fall somewhat short of the real test of performance.

And finally, personal communication is the highest-fidelity option for assessing oral communication skills. But, personal communication can only serve as a means to get at proficiency in describing sound performance: It falls short of providing direct data on the ability to perform.

### Assessing Products

If our assessment goal is to determine if students can create a certain kind of achievement-related product, the best way is to have them create one for evaluation. In fact, performance assessment represents the only means of direct assessment. The best test of the ability to throw a ceramic pot is the quality of the finished product. The best test of the ability to set up the proper science apparatus is the set up when it is completed. The best test of the ability to write a term paper is the finished paper.

Again, a selected response assessment format would reveal if a student can pick out the best-quality product from among several, as could a test of knowledge of the key attributes of a good-quality product. But these are limited substitutes for an assessment that actually asks students to create the product.

A student could write an essay or discuss the key attributes of a carefully crafted cabinet in shop. Again, the student who could respond well to such an exercise might possess the important prerequisite knowledge. But sometimes the real question is, "Can the student create a carefully crafted cabinet, an engine that runs smoothly, a high-quality term paper?" When these are the questions, performance assessment is the answer.

### Assessing Affect

Some of the affective dimensions of individuals that might be the object of classroom assessment include attitudes, values, interests, self-concept, and motivational dispositions.

The key to success in assessment is a clear and appropriate definition of the affective characteristic to be assessed. Given such a clear understanding, you could translate some such outcomes into selected response questions. But our collection of such items won't represent a test per se—they comprise a questionnaire. We could offer students statements and ask if they agree, disagree, or are undecided. Or we could ask them to select from among a list of adjectives the ones that most accurately apply to themselves or to some other object. And open-ended questionnaire items can be written in a way that asks students to describe both the direction and intensity of their feelings about any particular object. This assessment realm is rich with useful options.

The match between affective targets and performance assessment is a bit more complex, however, because of the nature of the inferences that must be drawn. We can look at a sample of student performance or at a student-created product and draw some conclusions about attitudes, values, and motivation with respect to that particular project. If students demonstrate high levels of achievement, they probably had a positive attitude, valued the project, and were disposed to work hard to perform well. However, a poor attitude and a lack of motivation is only one possible explanation for poor performance. Only additional follow up assessment will reveal the real reason for failure to perform.

One excellent way in which to conduct such a follow up might be direct personal communication with the student. In the right atmosphere, students will talk openly and honestly about the strength and direction of their feelings. The keys to success, of course, are to establish that open environment and to know what kinds of questions tap the important affective or social dimensions of the students.

## Summary: When Is Performance Assessment the Answer?

Performance assessment is the answer to our assessment needs in instances when the achievement target to be assessed involves student demonstration of complex reasoning, achievement-related skills, or competence in creating specified products. If our objective is to assess student mastery of content knowledge, proficiency with some simple reasoning processes,or attainment of particular affective outcomes, other assessment methods may be preferable. Further, if resources are limited, we might need to rely on other methods to approximate our most valued outcome.

Let me conclude with an example illustrating this latter point. When I teach my introductory classroom assessment course in the graduate program at Lewis and Clark College in Portland, Oregon, my goal is to provide teachers with the knowledge and skills needed to address real-world classroom assessment challenges successfully. At the end of the semester, the very best way for me to assess the competence of my students would be to take over a real classroom, set up circumstances to make specific assessment problems occur, and watch each of my students deal with those problems. Those who are competent will succeed. Another assessment option would be for me to follow them into their own classrooms to watch them solve real-world classroom assessment problems. If I observed with clear performance criteria in mind, I could assess well. Without doubt, this would represent the most authentic assessment I could possibly conduct.

At the same time however, we must acknowledge how unrealistic this assessment plan is. I have neither the opportunity nor the resources to conduct such assessments. Given this reality, how should I proceed?

My compromise is to use traditional essay assessment methodology as my best possible approximation. I present written descriptions of real-world classroom assessment challenges and ask my students to tell me in writing how they would solve the problem. If they are masters of the material covered in the course, they know how to use those tools and ideas to meet the challenge. If they are proficient at classroom assessment, that is clear in their written problem solutions. Admittedly, this is second best. But not too far from the real thing.

Over the past few decades of educational assessment in American schools, we have made the mistake of assuming that any outcome worth achieving can be translated into multiple-choice test items. And we ignored any that couldn't. That was an extremely counterproductive position that ignored the potential contribution of other available assessment options. Let's not make the same mistake again by letting that pendulum swing too far in the other direction in the 1990s. Let's keep performance assessment in perspective.

### REFERENCES

Bloom, B.S., M.B. Englehart, E.J. Furst, W.H. Hill, and O.R.Kratwohl. (1956). *Taxonomy of Educational Objectives: Handbook 1, Cognitive Domain.* New York, N.Y.: Longman.

Marzano, R.J. (1992). *A Different Kind of Classroom: Teaching with the Dimensions of Learning.* Alexandria, Va.: Association for Supervision and Curriculum Development.

Quellmalz, E. (1987). "Developing Reasoning Skills." In *Teaching Thinking Skills: Theory and Practice*, edited by J.B. Barron and R.J. Sternberg. New York, N.Y.: W.H. Freeman.

# 11

# *Placing Student Performance Assessment Within the Context of School Restructuring*

# II-1

# Placing Student Performance Assessment Within the Context of School Restructuring

## Introduction

H. Dickson Corbett
Research for Better Schools

Robert E. Blum
Northwest Regional Educational Laboratory

Schlechty (1990) argues that "restructuring means altering systems of rules, roles, and relationships so that schools can serve existing purposes more effectively or serve new purposes altogether." We believe that purposes to be served by restructuring must refer to *results for students*. The only compelling reason to start the arduous process of revamping the system is to change these results. So the restructuring process begins with determining what those results should be and then works backward to figure out how the system should operate to promote those results.

Student performance assessment becomes a consequence of these deliberations about students. That is, in a school or district that is truly restructuring, staff do not initiate the change effort to address assessment; rather, they begin with a clear and shared definition of what should happen with students. Then, later in the process, they address the kinds of assessments that can best inform them about students' progress toward these results. In their paper, **Corbett and Blum** describe restructuring as it is envisioned as a set of guiding principles for this book, using vignettes to illustrate how restructuring may proceed in a school district.

## Beginning with Results

Restructuring starts with developing a widely shared understanding about desirable results for students. This is no easy task. Key stakeholders—students, teachers, noncertificated staff, parents, administrators, nonparent community members, business representatives—must all think, learn, and determine together what students ought to achieve from the schooling process. Ideally, the focus would be what students need to know and be able to do to succeed in the 21st century. In reality, we need a lot of "scratching below the surface" of peoples' perspectives to make sure everyone is wedded to this focus.

That considerable controversy surrounds the issue of what students should learn is no secret. For example, is the purpose of schooling to teach reading, writing, and arithmetic? Or is the purpose to look beyond specific content acquisition to encourage the growth of reflective and caring lifelong learners? Even when people seem to agree on the surface, they may not, in fact, agree at all. Ongoing research in an urban elementary school, for example, shows that even though teachers use similar terms to describe the kinds of people they would like to see students become, they attach different meanings to those terms (Corbett, Wilson, and Webb 1996). Indeed, to speak of "the teachers' perspective" in this school would be misleading, because their views of desirable results for students actually fall into three categories. One group talks about what would make students successful in school; another describes characteristics that would enable students to be successful in life; and a third bridges both of these perspectives.

Interestingly, all three groups use the same descriptors: "self controlled" or "disciplined," "respectful," "confident," "responsible," and "skilled." The difference lies in the meanings attached to these words. The first group sees the characteristics as instrumental in making a school run more smoothly, and the second views the attributes as instrumental in the children's future outside a school environment. The third draws from both perspectives.

These below-the-surface differences are likely to result in contradictory actions that different staff members take in day-to-day school operations. These are the very contradictions that restructurers are seeking to avoid by altering the rules, roles, and relationships so that these elements align with desirable student results. Therefore, "beginning with results" is not to be taken lightly; it will probably be some time before a restructuring school or district is ready to turn to challenging issues of student performance assessment.

Whatever label is applied to this initial aspect of restructuring—such as developing a mission, creating a vision, or establishing student learning goals—this statement of purpose for students will become the foundation for whatever follows, and the effectiveness of subsequent actions will depend on how strong this foundation is. In other words, if we are to align actions with desired results, these results have to be clear and believed. Section III, Developing Student Learning Goals for the 21st Century, presents issues and processes on what students should learn. The section explores varying perspectives on what is important, approaches to defining student learning, who must be involved in establishing student learning goals, and decision-making criteria.

## Aligning the System with Results

Aligning what happens in school with student learning results is also difficult, but possible. It requires that those concerned with education continuously ask themselves, "What does—or what will—a particular practice or action do to ensure that all students learn well in all intended goal areas?" In other words, results become the criteria by which daily and long-range decisions are made. Asking certain questions helps focus on such results:

- Teachers must ask, "If I teach this lesson as I always have, will I be aiding learning in the desired goal areas for all my students?"
- Administrators must ask, "If I make the decision about budget cuts as I have in the past, will I be facilitating or hindering the staff and community in engaging all students in the kinds of opportunities they must have to achieve the results we want?"
- Parents must ask questions such as, "If I allow my child to stay home primarily because an assignment is not complete, am I helping my child develop the critical qualities that successful learners must develop?"
- We must all ask questions such as, "Which will result in the kind of learning students need—covering all the material in the textbook, or covering less material but covering it in a way that promotes critical thinking?"

A student-results direction set by the community and the district must guide day-to-day and long-range decisions. *All stakeholders must know about the results for students and allow these results to direct their actions.*

Aligning student learning results with stakeholder decisions and actions at all levels—district, state, school, and classroom—is critical. Support structures such as personnel, staff development, student assessment efforts, curriculum development work, district and state policy, and resource allocations must reflect this alignment. Everyone at the district and state levels should routinely use the agreed-upon student learning goals as the basis of decision making. They must question how their work will help move student learning toward these goals.

In her paper, **Barbara Benham Tye** sheds light on the meaning of the "deep structure" of schooling, adding to our insights about restructuring. According to Tye, widely held values and assumptions about schooling determine the deep structure. One common characteristic that makes up the deep structure of schooling is reliance on test scores as measures of success. Tye goes on to say that each school has a unique personality that is formed by a configuration of characteristics, such as how teacher, students, and principals work together. Changing the deep structure in a school along with changing its unique personality is difficult but essential. If both are not changed, restructuring is unlikely to affect student learning over the long run.

How does this deep structure relate to student performance assessment? Let's assume that a school staff, without much involvement from other stakeholders—parents, students, and community members, for example—decides to focus on a new, broad set of student learning goals. The goals include being able to work well in a team. The staff then begins to change, in part, how they assess student success, using performance measures rather than tests. Let's also assume that some students who have always done well on tests perform poorly in team learning activities. This situation sets the stage for a confrontation between some parents and staff, because the unique personality of the school has changed. Teachers are using new measures for student success, but a characteristic of the nationwide deep structure, parental assumption that success is measured by tests, has not changed. The school began a restructuring process to align assessment with student learning goals, but staff was not conscious of the deep structure characteristic related to the attempted change.

## The Systemic Approach

Using results-driven restructuring is not linear—it is systemic. When people are committed to using student learning results in agreed-upon goal areas as the criteria for decision making, they will begin to make decisions immediately. Individual teachers will begin to change lessons, and teams of teachers will integrate their work to make learning more meaningful for students. Students will ask questions of their teachers, schools, parents or other significant adults, and themselves, and they will expect assistance from everyone in their learning process. Parents will change how they interact with the school. And administrators will collect, organize, and use information differently in decision-making processes.

Some decisions will be long and painful, resulting in changes to the deep organizational structure. For example, deep structure characteristics such as schooling takes place in buildings called schools and student learning is assessed with paper-and-pencil tests may change. Other decisions will come easily, resulting in small but important changes in relationships among teachers, students, and parents. It will all be happening at once. For linear thinkers who have a high need for control, this systemic approach will be uncomfortable, and there is no easy remedy for this discomfort. Regardless of how restructuring begins or proceeds, the results must enhance the lives of all key stakeholders—students, staff, parents, and the community—if it is to be successful and lasting.

## Tying Performance Assessment to Restructuring

Discomfort from using a systemic approach may be especially acute for people involved in designing assessment procedures for a restructuring system. What is assessed signals what is ultimately important for students to learn. We argue that a system has to decide what is important for students to learn first, and then ideas about how to assess these learnings can be enacted. But the excerpt from *Education Daily* puts forth another point of view. It argues for making decisions about student learning by putting the assessment system in place first. Regardless of one's stance, an assessment system can be either a major positive force in promoting agreed-upon student results or a powerful negative force, producing confusion and mixed messages about results.

In his paper, **Fred M. Newmann** takes us back to beginning the process of restructuring with intended student learnings, providing additional insights into authentic achievement, and connecting what students learn in school to what people do throughout their lives. This paper shows the strong ties—perhaps even the unity—of achievement with assessment.

The paper by **Joseph P. McDonald** extends an understanding of the relationship between the results we seek and our assessment practices. An important point is that when we are clear about the intended results and have meaningful assessments to show progress, we can no longer ignore poor student performance over long periods of time.

### REFERENCES

Corbett, H.D., B.L. Wilson, and J.L. Webb. (1996). "Visible Differences and Unseen Commonalities: Viewing Students as the Connections Between Schools and Communities" in *Coordination Among Schools, Families and Communities: Prospects for Educational Reform.* Albany, N.Y.: SUNY Press.

Schlechty, P.C. (1990). *Schools for the 21st Century: Leadership Imperatives for Educational Reform.* San Francisco: Jossey-Bass.

# II-2

# Thinking Backwards to Move Forward

DICKSON CORBETT
RESEARCH FOR BETTER SCHOOLS

ROBERT BLUM
NORTHWEST REGIONAL EDUCATIONAL LABORATORY

First published in *Phi Delta Kappan* 74, 9 (May 1993): 690–694. Reprinted with permission from Phi Delta Kappa and the authors.

# Thinking Backwards To Move Forward

*Before "unstructuring" and "restructuring," a school or district or state should go through four steps. Messrs. Corbett and Blum provide the details.*

BY DICKSON CORBETT AND
ROBERT BLUM

THERE SHOULD be only one reason to restructure our school system: the students. We will take that premise as the starting point of our description of a process that should *precede* and inform any restructuring efforts. Before "unstructuring" (dismantling policies and practices that impede successful learning) and "restructuring" (creating new ways of operating that support successful learning), a school or district or state should go through the following four steps: 1) think first about students, 2) adopt successful learning as the criterion by which to judge the appropriateness of all daily actions, 3) expect all adults to be successful learners as well, and 4) identify the kinds of "rules, roles, and relationships" – that is, the social structure – needed to support successful learning.[1]

*DICKSON CORBETT is co-director of the Applied Research Project at Research for Better Schools, Philadelphia, Pa. ROBERT BLUM is co-director of the School, Community, and Professional Development Program, Northwest Regional Educational Laboratory, Portland, Ore.*

*The preparation of this article was supported by funds from the U.S. Department of Education, Office of Educational Research and Improvement. The views expressed are those of the authors.*

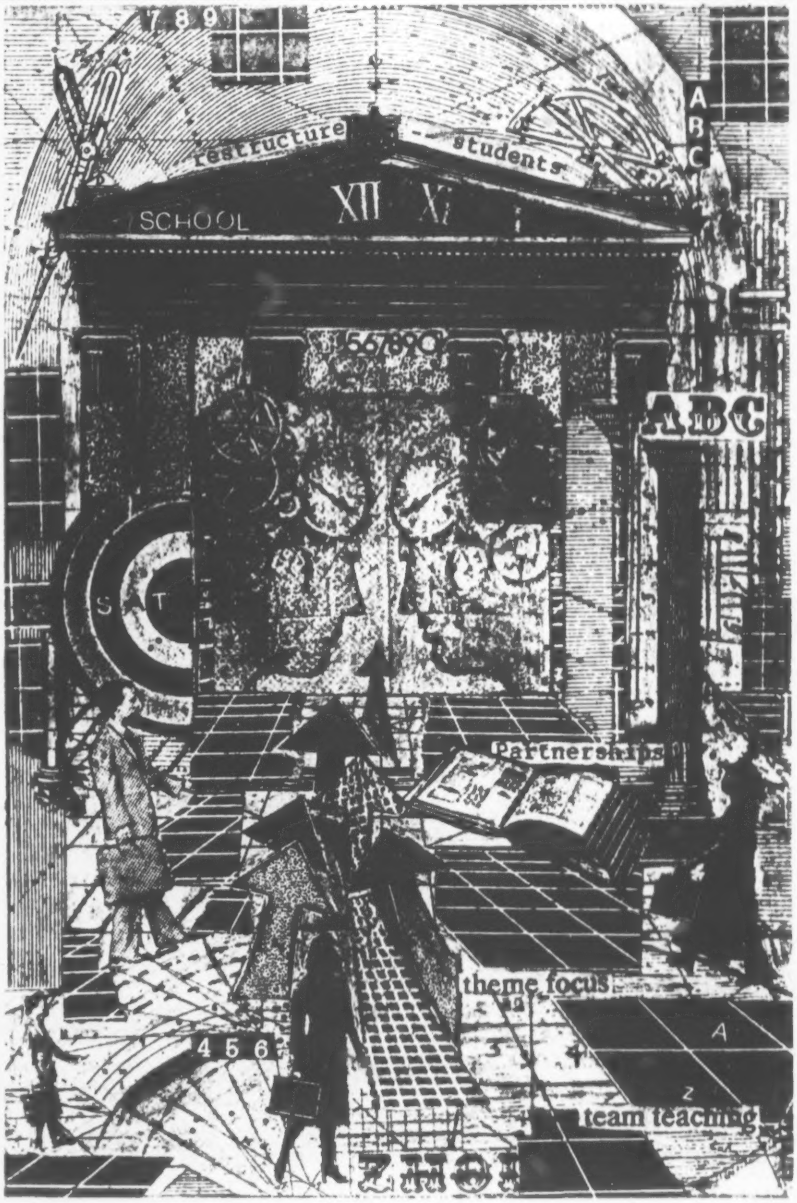


Illustration by Robert Mass

## THINK FIRST ABOUT STUDENTS

The rented van rumbled down the highway, its passengers lost in thought, rhythmically jostling one another. The six teachers, all from a large middle school located in the East Salem School District, had just left a two-day conference on cooperative learning.

Brenda broke the silence. "You know, every time I spoke about what we're doing at Knollwood, I realized that our kids are doing a lot of stuff, good stuff, that kids in other places – with a whole lot more going for them – don't seem to be doing." Noting the nodding heads, she continued, "And I don't mean just knowing things. They can make a schedule so that their projects get done on time . . ."

"Not all of them," interrupted Blair.

"No, of course not," replied Brenda, "but many of them can and do. And they can work together, and argue without making it personal, and look someone in the eye when they talk to them, and figure out ways to do things that I hadn't even considered."

"I know what you mean," said Steve. "They just seem a little more with it when it comes to acting like they'll have to when they go out into the world."

Dale picked up on Steve's thoughts. "Like when the Blue Team kids were upset about not seeing their friends from the other teams at lunch. A group of them asked for some time to meet to think about ways to rework the schedule. Then they had the whole team decide on the three best ones, and they presented the options to the faculty."

"A few years ago the only thing that would have happened would have been that a couple of kids would have complained a hundred times during class," said Debra.

"I don't want to throw a wet blanket on all of this. . . ."

"You always do, Brenda," David retorted good-naturedly.

"It's just that we in this car seem to agree that this kind of stuff's important, but we never get around to really saying what 'this kind of stuff' is."

"You mean like in a mission statement?" moaned Steve. The others joined in.

"That's exactly what I mean! What we're talking about is what will really make a difference when our students are running this world. We know it and believe it and act on it. But nobody else knows what we think, and we don't know what they think. What if everyone in our district agrees that self-motivation and caring for other people and creative solutions and knowing how to get information to come up with those solutions are what education should be about . . . and yet we never find out that we all agree? Our schools will just continue to have seventy-eleven objectives for each subject, teachers will mindlessly try to cover them as if those things are important, and we all will have to try to convince the kids that if they remember this stuff long enough they might get to use it one day! Come on! We know there's more to it than that!"

The road to restructuring begins with conversations like this one. While that road can ultimately lead to a lot of major changes – shared decision making, alternative assessments, new partnerships, flexible uses of time, and so on – the first step is to figure out *why* people should even make the journey. Brenda was pushing her colleagues to do just that. In actuality, similar conversations would have to take place throughout the school system and the school community – and over a considerable period of time – before Brenda's question about districtwide consensus could be answered. Key stakeholders need time to learn together, think together, plan together, and reach consensus on a vision of what students should become. In other words, schools have to be forward-thinking places where people are concerned about answering a central question: What characteristics do students need to begin to develop now in order to become responsive and responsible people in the future?

A review of the research on learning by Beau Fly Jones and Todd Fennimore begins to answer this question. Jones and Fennimore define a successful learner as one who is *knowledgeable* (able to locate and produce needed information), *empathetic* (able to appreciate and adopt multiple perspectives), *self-determined* (able to initiate, direct, and evaluate one's own learning), and *strategic* (able to solve problems creatively).[2]

However, just because people possess these individual qualities does not ensure that they will necessarily use them "out in the world" in a responsive and responsible way that will benefit others and themselves. Indeed, there is increasing concern that schools do not help students make connections between what they do in school and what they do outside of school and that students leave school without the ability to apply school learning to the activities in which they will engage throughout their lives.[3] Ernest Boyer contends, for example, that young people must develop as citizens and that the work of democracy is carried on through thoughtful discourse among people.[4] Others want students to succeed in the workplace, as family members, as aesthetic people, or as healthy individuals.

Encouraging learners to make the critical transition to real life requires attending to the social relationships that individuals construct. Recent work on the concept of caring seems best to focus ideas about learning in the context of those relationships.[5] In fact, caring – the ability to discern and act in the best interests of others – completes Jones and Fennimore's view of the successful learner by binding knowledge, empathy, self-determination, and strategic problem solving to the development of beneficial, reciprocal relationships through which the common good is pursued.

Different labels may be used in different communities to describe the kinds of students sought. For example, as Brenda and her fellow passengers in the van expand their conversation to include more people, they will probably begin to develop locally meaningful categories of characteristics that they want to see students develop. Whatever labels are used, collectively they need to capture all aspects of successful learning and its application in the day-to-day world. In other words, student outcomes must be both *locally determined* and *soundly based*.

## USE SUCCESSFUL LEARNING AS THE CRITERION FOR ACTIONS

People from across the district settled into their places, looking one last time at the reports they would give to one another about their schools' progress toward aligning the curriculum with the mission. Brenda reflected on all the discussions about students and schooling that had occurred in the district during the previous year, involving colleagues from other schools, the district office, and the community. The year had been both frustrating and exciting – frustrating in that nine months seemed like a terribly long time to spend just talking, but exciting in that people had begun to see that there was a common core of understandings about what students should be able to do. Those understandings had been formalized in a mission statement, which expressed a much broader view of the purpose of schooling than had ever been publicly acknowledged:

> The East Salem School District believes that all students can be successful throughout their schooling experience. Our purpose is to create a nurturing environment that will foster knowledgeable, empathetic, strategic, self-directed, and physically fit individuals who will be able to function successfully in society. When students complete their schooling experience, they will be ready to enter the next stage of their journey to become productive citizens.

Brenda was relieved that all of the discussion had made people realize that they should probably rethink their approach to teaching and learning. They began to understand that it would be very difficult to achieve their mission through the schools' current subject-oriented organization. The discussions had reminded them that the world is not organized in discrete content areas and had opened the door to thinking about new ways to organize curriculum, instruction, and assessment. Vast changes were probably needed.

At Knollwood, the faculty and the principal agreed that they would begin by reexamining their subject-based approach. They decided that the initial task – and the one that Brenda was preparing herself to discuss in the meeting – was to put teachers from different teams and different subject areas together and have them identify core student outcomes that cut across content areas, keeping the larger mission in mind. The hope was that this task would help people begin to see beyond content-specific objectives to overall learning goals.

Brenda knew that not everyone found this easy. Still, it hadn't taken teachers long to come up with several outcomes for each of the academic subjects. They all agreed that "knowing the process of inquiry, being oriented toward problem solving, and knowing where and how to look for information" were important. They were less sure about the fit between academic content and some of the more "affective" outcomes. Brenda knew that, because of this difficulty, something significant was missing from their recommendations. "Because so much of our mission concerns how people should relate to one another," she reported to those at the meeting, "I wanted a social outcome in there for math, English, social studies, and science. I was pushing for these subjects to match the mission . . . but I didn't get it."

Reaching a widely shared understanding about successful learning may be the starting point of restructuring, but this understanding pretty quickly has to become explicitly embedded in day-to-day action. A vision of what a community wants for its students establishes the basis on which all subsequent restructuring decisions should be made. It helps to sort out what is "good to do" from what "should be done."

Once the Pandora's box of restructuring has been opened, a district can more easily cope with the myriad demands for special programs and policies if it has already set the criterion by which the worthiness of such demands will be judged. Will this program or policy in some way "enable students to develop the characteristics they need to be successful, contributing members of the community and society in general?" Ultimately, the mission's focus on successful learning will have to propel every interaction, every activity, and every purposeful endeavor that takes place. Musical programs, curriculum decisions, sports programs, field trips, assessment techniques, budgeting procedures, schedules, staff development initiatives, decision-making patterns – indeed, any change at all in the structure of schooling – will have to meet this one criterion. That is why it is imperative that a school community's view of successful learning be both shared and clear, so that it can be consciously and concretely applied to the ordinary routines of school.

The test of whether a particular value or perspective has been internalized lies in having it emerge naturally in people's daily interactions. Brenda's group had not reached this point yet. For their mission statement to become a working document, they would have to think differently about subject areas and about how to organize themselves for teaching and learning. A continued failure to orient teaching and learning to the student outcomes would eventually render the mission hollow.

### EXPECT THE ADULTS TO BE SUCCESSFUL LEARNERS

Parents, teachers, the custodian, and the superintendent wandered through the classroom – if you could call it that. The typical accouterments weren't visible. In their places were animals: dogs, cats, elephants, snakes, rabbits, lambs, giraffes, bears, even a manatee, each in its own colorful habitat, comfortable, surrounded by all the materials it liked to nestle in and the foods it liked to eat. Scattered throughout the displays were concession stands where people could "buy" brownies, cookies, or even little bags of food to feed the animals. Money smoothly changed hands, with each exchange recorded in a ledger. Of course, the money was only play and the inhabitants of this zoo were really stuffed.

Several students hovered around the animals and concession stands, eager for visitors to ask questions about how the real-life models for the toy versions lived, or how the students had constructed the habitats, or how hard they had worked together to build a zoo in two 40' x 40' rooms, or how they had figured out how much they had to charge for admission or for the items for sale.

The superintendent walked over to Brenda. "This is marvelous!"

"Even if we spent the entire month of March doing the research, constructing the habitats, and putting it together? Even if we suspended the regular math, science, and English lessons that whole time? Even if the kids didn't have homework?" Brenda inquired. She always wanted to know where things stood.

The superintendent laughed. "Yes, yes, and yes. I'd have to be pretty detached from the schools not to appreciate what the kids have accomplished here: participating in groups, sharing ideas, connecting school subjects to the real world, figuring out how to find information and how to apply it in specific situations, thinking critically and creatively, reflecting on their experience for all these visitors. This stuff is cooperative learning, higher-order thinking, and authentic assessment all come alive!"

"What if every teacher in the district did things this way?"

"Well, we'd have to look even more seriously at curriculum and testing and all that. But you don't really think everyone should do a zoo, do you? Don't we have to allow teachers to teach in the way they feel comfortable? They all have different strengths and weaknesses. I firmly believe that diversity is important."

"Yes it is, and not everyone should do a zoo. However, don't you think that the things you just listed are things that every child should have the opportunity to accomplish? Comfortable teachers are okay; diversity is fine. But what if what teachers are comfortable with sells the kids short? What if diversity makes it so that some kids end up getting chances to do some of these things, some have chances to do them all, and some never have the chance?"

"Those are big questions. A lot of people would have to change in a lot of ways."

"Exactly."

Participating in groups, sharing ideas, connecting school subjects to the real world, finding information and applying it to specific situations, thinking critically and creatively, reflecting on experiences – these are all appropriate actions for young people to engage in as they become the kind of successful learners identified in East Salem's mission statement. Brenda's argument draws attention to some implications these desirable student results have for the adults in the system. If staff members (and parents) are to facilitate the development of students with particular characteristics, then they must acquire and model these characteristics themselves. The adults must know what it feels like to be knowledgeable, empathetic, self-determined, strategic, and caring. They must know what it is like to apply what they are teaching to out-of-school situations, and they must work consciously to help students make those connections. In other words, the only way that a mission based on successful learning can emerge as the sole criterion by which day-to-day decisions get made is if all the people making those decisions are explicitly and actively engaged in successful learning themselves. To use the superintendent's word, everyone has to become "comfortable" with the mission of successful learning before it becomes the heart and soul of school activities.

In the majority of school systems with which we have experience, there is a hit-or-miss character to the quality of education. The programs and activities that determine the extent of students' opportunities to become successful learners vary between schools and even between classrooms in the same school. The consequence is that all students may develop some of the intended characteristics, but only some students will be fortunate enough to follow a route through the system that encourages the development of all of the outcomes. A school district has to ensure that idiosyncratic variations in programs, people, and policies do not result in systematic differences in the quality of education children receive.

For example, students with minimal basic skills should not be relegated to a minimalist education devoid of attractive learning opportunities;[6] "average" students should have the same richness of educational experiences as the gifted; students with a first-year teacher should have no less, and no more, access to engaging activities than students with a veteran teacher; and a student with one fourth-grade teacher should neither be advantaged nor disadvantaged in comparison to a student with another fourth-grade teacher. Thus the benchmark of a school or district trying to encourage successful learning in students is an active commitment to successful learning on the part of all adults, including support personnel, parents, and other community members. The adults serve as models of the qualities expected of students and are skilled in ways of establishing activities that promote these qualities.

---

### IDENTIFY RULES, ROLES, AND RELATIONSHIPS

The third speaker stood at the podium. The previous two, a policy researcher and a state superintendent, had gone on in detail about how the state's switch from a mode of enforcement to one of assistance was the catalyst for significant changes in local districts. Despite the promise that this development held for education, the East Salem teachers in attendance were a little antsy. They had heard grand designs before; what they wanted to hear about were concrete positive events that actually affected students. With anticipation they edged forward in their seats a bit as the curriculum director of a nearby district began to speak.

"I want to share with you all of the exciting things we've been doing for the last two years and to give you an idea of how we got where we are today. It all started with strategic planning. Small groups of parents, teachers, administrators, and community leaders met for several months to help us sort out where we wanted to go. . . ."

Brenda leaned over to David and whispered, "What about the students?"

"Of course we realized that we needed to rethink how we used our time. Significant change couldn't be made if the teachers didn't have time to talk to one another. So with teaming, rechunking time periods, and blocking special classes, we managed to find a way that people could meet for an hour every day to work on our restructuring committees. . . ."

"What about the students?"

"But we realized that, without the commitment of everyone, none of these scheduling changes would foment the kinds of interaction we wanted among the professional staff. That brought us to shared decision making. . . ."

"Where are the students?"

"Brenda, quiet! You sound like a Wendy's commercial!"

"And due to all this hard work, we have in place a system that enables us to achieve our mission – 'A Quality Education for All.' "

Later, as the audience filed out of the room, the Knollwood contingent discussed what they had heard.

"So, Debra, what did you think?"

"I agree with Brenda. It sounds to me

*"Fluffy is enrolled in an advanced education program."*

like a bunch of people have put the cart before the horse. They adopt all these new ideas without really knowing what the purpose is, other than to shake things up or look innovative."

Blair chimed in. "Those people have made all of these changes, but in an hour and a half they never said anything about what the students are doing differently. It's like they think that if all the adults are happy, then somehow magically students will be better off. . . ."

"Without ever saying what *better off* is!" interjected Dale.

"Other than making a big deal about little gains on some silly old test," Debra added.

Steve started to say, "Well, as Brenda's been saying. . . ."

"Over and over and over again!"

"As I've been saying, David, it's been two years since the six of us had that conversation in the van. And now, because we and the other teachers at Knollwood and people throughout the district have been talking to one another – whenever there's a chance – about what we want to have happen with students, it's like we're all in the van together, and now we're ready to make some changes that are connected to students."

"Finally!" David exclaimed. "I say it's about time!"

"It *is* about time. But two years ago was *not* the right time, because then we didn't know what we wanted to accomplish."

"So now if we decide, say, to move into alternative assessment, we'll do it because we think that, in some way, it will encourage teachers to create the kinds of activities students need to do to be knowledge producers and self-motivated learners and all that."

Brenda concurred with Blair. "That's right. Or, if we decide teachers should be guides rather than dispensers of information, it's because, considering our particular mission, being a guide is a better way for teachers to go about teaching."

The conversation continued about possible changes and which ones would most likely be effective in promoting East Salem's definition of success. Finally, Dale wondered, "What would have happened if we had done it the other way and just jumped into these things?"

"We'd be on to something else by now," concluded Brenda.

Partnerships, site-based management, alternative assessment, flexible scheduling – these are not the beginning points of the restructuring process. They should not be the stimulants of reform. Rather, they should be the *consequences* of extended dialogue about students, about daily activities, and about adults' professional commitments. These popular approaches are simply possible answers to the question, Given what a school community wants to see its students become, how can it operate so as to encourage that? That is, how do the current rules, roles, and relationships have to change in order to produce the results a district wants to see?

For example, if East Salem decided to pursue for all students the kinds of results Brenda was getting with her zoo activity, then the rules for organizing the curriculum would have to change. In fact, learning and teaching activities could no longer remain subordinate to subject areas in such a situation; instead subject-oriented skills and knowledge would be integrated into a curriculum made up of series of experiences that closely approximated out-of-school activities undertaken by youths and adults. Obviously a curriculum of this sort would also entail a redefinition of roles. Students would have to become very active participants in classroom, school, and community life. Indeed, sitting at their desks would become a welcome respite. Relationships would have to change as well, in terms of who interacts with whom, what they interact about, and the form the interactions take. Students would talk much more with other students about substantive matters, teachers would work intensively with individuals and small groups of students, and students would run into people outside the school walls more often.

It is worth noting that a major part of restructuring may well be deleting existing practices that clearly inhibit successful learning. Too often, new ideas just get dumped on top of the residue of old ones. The result is a melange of practices that lack focus or purpose. So, the first actual step in implementing restructuring might be *un*structuring – that is, taking apart the infrastructure of expectations, work styles, and interactions that supported the old way of doing business.

Even when unstructuring precedes restructuring, notice that the thinking process does not start with these changes; it ends with them in order to ensure that the system will not waste time pursuing approaches that do not improve the chance that all students will develop the agreed-upon characteristics. Having thought backwards from student outcomes to structure, people in a school community are then ready to move forward with the actual changes. At this point, "It's finally time to do something," as an exasperated and relieved David said.

The brief glimpses of school life contained in the above vignettes should make it abundantly apparent that 1) restructuring takes a lot of time, energy, and courage on the part of everyone and 2) restructuring is an inherently messy process, full of talking, thinking, acting, sharing, and adjusting. Quite simply, if substantial, systemic change is to occur, all key groups of stakeholders must be engaged in a very time-consuming process. Consequently, there is a need for a lot of learning – about the substance of the learning process itself, about the kinds of changes that have been attempted elsewhere, and about the values that participants bring to the conversation. Thinking backwards is a way of trying to ensure that all of this effort will be both meaningful and beneficial for the entire school community. Equally important is the view of student results from which a group is thinking backwards. While this view must be determined locally, it must also be representative of what successful learning truly is, in both its individual and its relational aspects. Otherwise, moving forward will prove to be a detrimental rather than a beneficial step for society as a whole.

---

1. Everett Wilson, *Sociology: Rules, Roles, and Relationships* (Homewood, Ill.: Dorsey, 1971).
2. Beau Fly Jones and Todd Fennimore, *The New Definition of Learning* (Oakbrook, Ill.: North Central Regional Educational Laboratory, 1990).
3. Lauren B. Resnick, "Learning In School and Out," *Educational Researcher*, December 1987, pp. 13-20.
4. Ernest L. Boyer, "Civic Education for Responsible Citizens," *Educational Leadership*, November 1990, pp. 4-7.
5. Alfie Kohn, "Caring Kids: The Role of the Schools," *Phi Delta Kappan*, March 1991, pp. 496-506, and Nel Noddings, *The Challenge to Care* (New York: Teachers College Press, 1992).
6. Jeannie Oakes, *Multiplying Inequalities* (Santa Monica, Calif.: RAND Corporation, 1990). K

# II-3

# The Deep Structure of Schooling

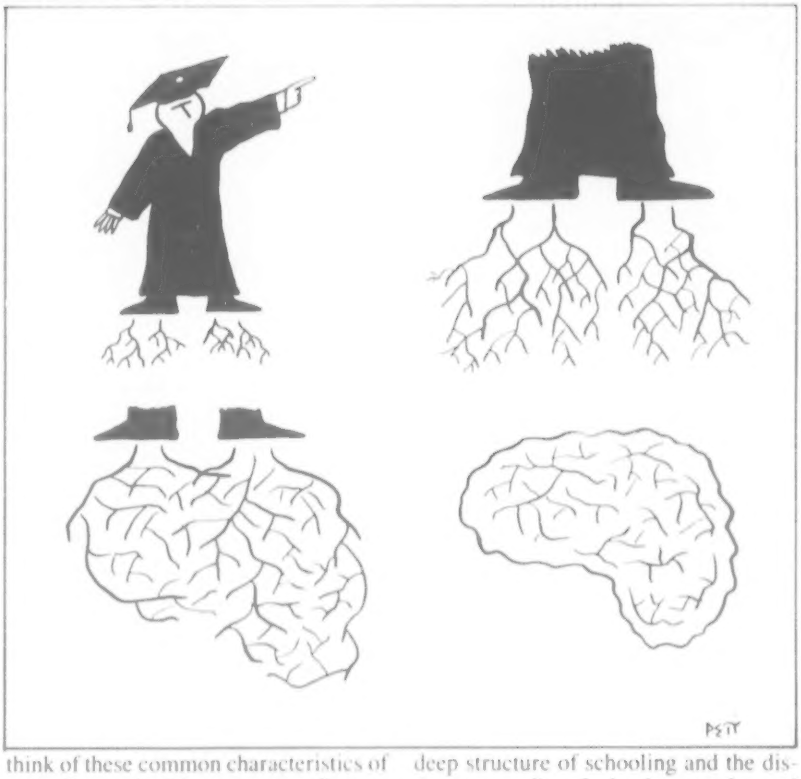
Barbara Benham Tye
Chapman University

First published in *Phi Delta Kappan* 69, 4 (December 1987): 281–284. Reprinted with permission from Phi Delta Kappa and the author.

# The Deep Structure Of Schooling

*No matter where you go in the United States, public high schools have a great many things in common. Ms. Tye examines this "deep structure" and explains how it affects any efforts to change the way schools work.*

BY BARBARA BENHAM TYE



AMERICAN HIGH schools are very much alike. But at the same time they are very different. This is one of the persistent paradoxes of American schooling. The findings of John Goodlad's Study of Schooling reaffirm this paradox.[1] In terms of what Seymour Sarason has called the "regularities of schooling," all 13 high schools that Goodlad's team studied were very much alike indeed.[2] This was true even though they varied in size from 64 to 3,006 students, in location from inner cities to isolated rural settings, and in socioeconomic level from poor to upper-middle-class.

Walk into a public high school in any of the 50 states, and you are likely to find yourself in familiar territory. You will not be surprised by the physical uniformity of classrooms; the overall control orientation of policy, program, and pedagogy; the general similarity of curriculum and of schedule; the reliance on test scores as measures of "success"; and the practice of tracking. I have come to think of these common characteristics of schooling as its "deep structure."

Yet each school I studied as part of the Goodlad team was also different from the others in dozens of big and little ways. The cumulative effect of these differences gave each school its particular "personality." Each of the 13 high schools was shaped by its own history, by the nature of the community of which it was a part, and by such internal factors as the quality of teacher/administrator relationships, the number and intensity of school problems, and the climate of most of its classrooms.

This juxtaposition of concepts — the deep structure of schooling and the distinct personality of schools — can be used heuristically to think about the problems of change and resistance to change in our educational system. Some readers might be inclined to argue that the similarities of schools obviously outweigh their differences, since the likenesses have to do with the deepest structures and assumptions of schooling in our culture. As long as one is primarily concerned with *describing* schools, it may be true that commonalities are more important.

However, if we are concerned with the processes involved in *improving* schools, the differences become equally impor-

---

*BARBARA BENHAM TYE is an associate professor of education at Chapman College, Orange, Calif.*

Illustration by Joel Pett

tant. The key to changing a school lies in the particular configuration of forces that give *that* school its unique personality. For this reason, a change strategy that works at one school might not work at another, and changes mandated at the district or state level will be implemented with varying degrees of success in the schools to which those mandates apply.

---

## DEEP STRUCTURE

The deep structure of schooling is determined by the values and assumptions that are widely shared throughout our society. Americans do not vary greatly in their views of desirable and appropriate educational experiences for children and young people. These values and assumptions are also shaped by conventional wisdom, by tradition, by vested interests – and by a certain amount of institutional inertia.

I have already mentioned several components of the deep structure. Let us examine one in particular. The assumptions that undergird decisions about the use of space in schools form one component of the deep structure. Whether you spend your days in an "old-fashioned" three-story school building or a "modern" one-story finger- or pod-style building, the classrooms are awfully similar. Whatever the design of the school, the physical environment of the high school classroom reinforces the passive behavior of students and makes it easier for adults to contain and control them.

First, all of the 525 high school classrooms we studied in A Study of Schooling were large enough to accommodate between 20 and 35 people, but *only if those people were behaving in certain ways*. The standard high school classroom is not large enough for 30 people who are square dancing, fencing, or practicing karate. More realistically, it is often not even large enough for six small groups to carry on discussions without disturbing one another. The typical high school classroom is just about right, however, for 30 people to watch a film, listen to a speaker, take a test, or read a book. This is true because all the latter activities require that people engaging in them be *seated* and *noninteractive*.

Second, whether the room is furnished with individual desks or with tables large enough for six people, there is a place for every person to *sit down*. Not a cushion or a couch, mind you, but a chair. Furthermore, students are often assigned to a *certain* chair and told that they may use no other. Moving around the room is discouraged, and the range of possible learning activities is further narrowed to those that can be done while seated in a chair.

Finally, the desks or tables in the typical high school classroom are usually arranged in rows, facing in one direction. This arrangement reduces the amount of interaction possible among students. Frontal presentation to the whole group was the teaching method most often used by the teachers in our sample. Was this because the desks were all facing the front of the room? Yet the desks were almost always *movable*; the teacher *could* easily enough have used a nonfrontal method, such as small groups. In the past, when desks were customarily bolted to the floor, teachers may have been required to lecture more than they might have liked; that is no longer the case. Today, however, form follows function: the seats all face the same way *so that* the teacher can lecture.

The rationale for this heavy reliance on frontal teaching is that the teacher possesses certain knowledge that the students need. Because the students are at a given grade and ability level, the teacher assumes that almost all of them need to learn the same things. True, a few probably know the information already, and others may never learn it, but *most* of them need to learn it and can learn it. The teacher must, therefore, reach the majority. Besides, if a teacher is talking to the whole class at once and can see all their faces, *it's a lot easier to stay in control* and keep them all quiet.

Thus we see a strong connection between the use of space and what I have termed the "control orientation" of schooling. Both are aspects of the deep structure – so taken for granted that we almost never think to challenge them. An emphasis on controlling students has been part of the deep structure of American schooling since colonial times, when society regarded children as little animals in need of taming. This legacy is still with us, and it determines the decisions of school boards and architects, as they continue to design high schools full of the standard classrooms I have described.

The control orientation of schooling is with us, as well, in the assumptions of teachers in the newer open- or flexible-space schools, in which many teachers yearn for traditional, closed-off classrooms and plead for the construction of permanent walls. And it is with us in the assumptions of principals, parents, teachers – in fact, most of us – that a quiet room is a room in which learning is taking place and a noisy room is a room in which nothing constructive can possibly be happening.

The control orientation of schooling can be seen in many other aspects of the way high schools are run. Closed-campus policies, for example, stem from the assumption that students *will certainly* abuse freedom if given the chance. Unfortunately, by removing the opportunity to misbehave, the school administration does not allow students the opportunity to behave well. A similar point can be made about disciplinary rules that specify what the teacher expects from the class. Yet one rarely finds any public acknowledgement of what the students have a right to expect from the teacher.

The custodial role of the American school is well understood by all of us, but it is seldom discussed. We want our young people to be safe, off the streets. As parents, we want to know where they are – at least for part of the day. This is an example of what I mean when I say that the deep structure of schooling is determined by values and assumptions that are widely shared throughout society.

It is important to remember this pervasive, *shared* aspect of the deep struc-

SUBTRACTION TEST
C-
Campbell

*"I can put 2 and 2 together, but I can't get them apart."*

**Improvement at the level of a single school's unique personality is unlikely to alter the existing regularities of schooling at a deeper level.**

ture. Each school does *not* have its own deep structure; the deep structure of American schooling is *nationwide*. It is this connectedness that gives the deep structure its persistence and its power.

### UNIQUE PERSONALITY

Built on the foundation of the deep structure, however, are the unique personalities of each school. At *this* level, no two schools are exactly alike. I have suggested that a school's power to change is determined by the configuration of characteristics that make up its unique personality.

For example, suppose that a school staff decides to work together to improve the quality of interaction among staff members – a characteristic that is part of a school's distinct personality. The teachers decide that they want to increase cooperation and sharing among themselves. Whether this goal is relatively easy or relatively difficult to achieve depends in part on the quality of the communication that already exists among the teachers; on the intensity of other school problems that also claim the teachers' time, attention, and energy; and on the nature of the relationship between the teachers and the principal.

In a school in which teaching materials and supplies are scarce and in which communication among teachers is poor, cooperation and sharing may be very difficult to achieve. Hoarding is more apt to be the norm. However, in a school in which there is healthy, open communication and in which the principal actively works with teachers to help them identify and solve problems, the chances of dealing constructively with the problem of hoarding and of improving cooperation and sharing are much better. This is what I mean when I say that a school's *power* to change is determined by the configuration of characteristics that make up its individual personality.

### CHANGE AND THE DEEP STRUCTURE

Goals for change can be set at either level. However, a change in one or more aspects of a school's unique personality may well leave the deep structure unchanged. For example, teachers may work together to improve the quality of interaction among staff members, but this will not necessarily affect the curriculum, the bell schedule, the use of frontal teaching, or the assumption that students must be kept under control.

On the other hand, a change in *any* aspect of the deep structure will always affect many aspects of a school's unique personality, as well. It is as if one must go *through* the level of the unique personality in order to reach the deep structure; in fact, it may be necessary to *postpone* change efforts at the deeper level until certain changes at the more accessible level have been achieved.

In proposing that changes at the level of a school's individual personality probably need to precede changes at a deeper level, I do not mean to imply that changes in the deep structure are more worthwhile or important, only that they are far more difficult to achieve. They disturb the equilibrium of a school far more profoundly, and, because the educational system is so complex, they run a much greater risk of failure. The elements of schooling that I have called deep structure are rooted in the values and assumptions of our society, they are part of the conventional wisdom about schooling, and they have come to be accepted without question.

It is unlikely, therefore, that staff members would consider tackling a fundamental change in a school's deep structure. If they *did* decide to do so, they would have their work cut out for them. Let's say that the teachers at a medium-sized high school decided to work on using a wider variety of learning activities and materials in their teaching. In effect, they agreed to try and get away from the dominant pattern of lecture/question/test. Even reaching agreement on this goal would be difficult, given the deeply entrenched norm of classroom autonomy that exists in American schooling. But suppose that the teachers did set such a goal for themselves. What barriers to change would influence their efforts?

To some extent, the teachers would have to overcome their own training and experience, which prepared them primarily to lecture. They would have to be emotionally willing to take some risks, by trying unfamiliar activities and involving students in new ways in their classrooms. They would also be likely to encounter resistance from some students and from parents who believe that lecturing is the "right way" to teach because that's how *they* were taught. The teachers might have to conduct an inventory of all the learning materials in the school and set up a system to insure that the materials would be readily available to those teachers who plan to use them. Perhaps the teachers would have to work with the principal, the parent/teacher association, or a district committee to locate funding for a wider variety of learning materials (newspapers, films, computers, etc.). They would also need to find consultants who could provide the inservice training that they would need in order to learn and practice the new teaching methods and learning activities.

These are just a *few* of the problems that would need to be solved in the course of a single effort to make a major change. On the whole, it is much easier to give up and return to the familiar, tried-and-true pattern: teachers talk; students listen and then take tests.

In my experience, most change efforts begin to falter in about the third year of implementation. A change that cannot survive the "Year 3 Crisis" is apt to have been a change that was undertaken without sufficient attention to all the systemic pressures and consequences involved.

Such changes as do take place in high schools tend to be at the more accessible level of the school's unique personality. Improving staff interaction, solving specific school problems, building mechanisms for shared decision making, and all the many other possible goals for school improvement at this level are extremely worthwhile. They can go a long way toward making an individual school a happier and more productive place. And that is all to the good.

However, American secondary education *as an institution* will not improve significantly as long as changes are made only at this level and only in individual schools. Changes in the deep structure of schooling, which is shared by all schools, must also occur, and these changes require the kind of substantive support that can come only from policy changes, accompanied by significant shifts in public perceptions of what schools should be and do.

For schoolpeople, thinking in terms of these two levels can teach us a number of important lessons about our own experiences with school improvement efforts. We can see, for example, that, without certain kinds of support, some battles may be virtually unwinnable. We would therefore be wise to choose our battles with care.

We can understand, too, that improvement at the level of a single school's unique personality is unlikely to alter the existing regularities of schooling at a deeper level. Indeed, the nature of the deep structure of schooling may thwart some change efforts at the level of unique personality. When we wish to make a change in our own school, we are often told, "Better not rock the boat." That's the voice of the deep structure we hear.

Finally, we can recognize why district or state mandates usually produce only superficial and uneven success: the idiosyncratic nature of the unique personality of each school thwarts efforts by educational leaders to institutionalize school programs in any sort of unified way throughout a district or state. Even when (as usually happens) teachers and administrators act in good faith to implement these mandated changes, the unique characteristics of a given school's personality can get in the way. A wise policy maker understands this principle and makes room for local modification of required programs.

---

1. See John I. Goodlad, *A Place Called School* (New York: McGraw-Hill, 1984); Barbara Benham Tye, *Multiple Realities: A Study of 13 American High Schools* (Lanham, Md.: University Press of America, 1985); and Kenneth Tye, *The Junior High: School in Search of a Mission* (Lanham, Md.: University Press of America, 1985). The findings of A Study of Schooling have also been reported in numerous articles and technical reports.

2. I am indebted to Seymour Sarason for the concept of *regularities*, both behavioral and programmatic. See Seymour Sarason, *The Culture of the School and the Problem of Change* (Boston: Allyn & Bacon, 1982).

# II-4

# Arizona Uses High-Skills Tests to Drive Reform in Schools

EDUCATION DAILY

While Congress debates whether to require states to have academic standards and related assessments, Arizona's school leaders are saying their experience shows that is the best possible path to school reform.

"It's the most powerful thing we can do," said Diane Bishop, the state's school superintendent. "It's actually making things happen."

Last spring, Arizona completed a 10-year effort on standards and assessments when it tested every third-, eighth- and 12th-grader. The tests will be given annually from now on, with each schools' results reported.

While it took seven years and two tries to establish curriculum standards, state officials got their testing program up and running in less than three years.

## Can Do Under Limitations

Arizona's experience putting a statewide test in place so quickly with a limited budget is a lesson for other states, one testing expert said.

"They certainly demonstrated that a state can develop a performance-based program within a short period of time," said Ramsay Seldon, director of state education assessment programs for the Council of Chief State School Officers.

To pay the start-up fee, Bishop said, Arizona cut back on its other standardized testing, using funds saved from that $1 million budget to pay for development and higher grading costs. It also cut costs by allowing Riverside Corp., the test designer, to retain a copyright and the ability to sell the tests to other schools.

While Arizona has its standards and assessment system in place, in Washington, D.C., the House and Senate differ as to whether to force other states to undergo similar processes to qualify for federal school reform grants.

Under H.R. 1804, the House's version of President Clinton's Goals 2000: Educate America Act, states would have lay out their plan to develop standards and tests to receive a fiscal 1994 program grant (ED Special Supplement, Oct. 6).

---

Reprinted from *Education Daily,* Vol. 26, No. 247 (Tuesday, December 28, 1993) with permission of Education Daily, Capitol Publications Inc., P.O. Box 1453, Alexandria, Va. 22313-2053 (telephone 703-683-4100)

The bill the Senate plans to take up early next year, S. 1150, would allow states to work on such standards and assessments while getting a piece of the $105 million appropriated for fiscal 1994 (ED, Nov. 19).

Both bills would form a federal panel to review and certify standards and assessment plans states would voluntarily submit.

Bishop said she would welcome that opportunity, adding that in Arizona, "I think we have a product that will pass muster."

### Hands-On Testing

The Arizona Student Assessment Program (ASAP) doesn't have the multiple-choice, fill-in-the-blanks or other standardized questions found on most tests. It requires students to respond over the course of several days to "real life" questions by using writing and mathematical skills.

The students "have to show their work and the process they used to come up with their answer," Bishop said.

One sample test requires students to understand biological and mathematical concepts to answer a set of questions about whether biologists should attempt to bring back to Arizona the Mexican wolf, which fled as the state was settled by ranchers who killed wolves that attacked their herds.

On the test's first day, third-graders read a letter from a state wildlife official explaining the deliberations over whether to bring the wolf back. The letter concludes by encouraging students to respond with their opinion.

The students are asked general questions about the letter, such as "What does the endangered species coordinator want you to do with the information enclosed with the letter?" and "What kind of reaction do you think the endangered species coordinator hopes to get from his letter?"

On the second day, the students read in-depth information about the wolves' characteristics, habitat and diet. They then are asked to write a letter explaining their beliefs about whether the wolves should be brought back to Arizona.

Later during a group discussion, the students must explain the basic issues surrounding the wolf plan and what additional information they want to know.

On the third day, students take data about the wolves, the areas the state is considering for their placement and other issues, and then they answer math questions. For example, students are given population figures for elk, deer and other prey and asked to compute which area would offer wolves the best food supply.

On the final day, the students must write an essay explaining their opinion about the wolves' reintroduction. They must complete checklists on the information they should include in the essay, and are encouraged to revise their essay's drafts.

"It requires writing and demonstration of their knowledge," Bishop said. "It asks: 'Do you understand the issues?'"

## Training Needs

Until Arizona administered ASAP and issued a school-by-school report of its results, it saw little movement in schools applying the curriculum standards the state spent seven years developing.

Now, with the specter of being labeled a public failure, schools are scrambling to change their curriculum to match the standards.

"For the first time, they realize these skills have to be taught because they are going to be assessed" every year, Bishop said.

## Plays In Peoria

The Peoria Unified School District in suburban Phoenix has arranged for its schools to close a half day each month so teachers may learn more about the standards.

"You see high school teachers who are designing courses...that will lead to [students'] success on these new standards," said Paul Koehler, district assistant superintendent for academic services and a state associate superintendent who helped design ASAP. "That's exactly what needs to be going on."

"Without [testing] data, many teachers may not have been convinced that the work had to be done," Koehler said.

In fact, when Arizona applies for a Goals 2000 grant, Bishop said, the state will propose "massive staff development" that would reach the state's 33,000 teachers.

"They are going to change what they're doing," she said. "There's a lot of communication, a lot of work that has to go on to reach each one of those teachers."

# II-5

# Linking Restructuring to Authentic Student Achievement

Fred M. Newmann
Wisconsin Center for Educational Research
on Organization and Restructuring of Schools

First published in *Phi Delta Kappan* 72, 6 (February 1991): 458–463. Reprinted with permission from Phi Delta Kappa and the author.

# Linking Restructuring to Authentic Student Achievement

*Unless plans for new structures are guided by a vision of educational outcomes that articulates new standards for student achievement and for teaching, restructuring is not apt to produce favorable results for students, Mr. Newmann warns.*

BY FRED M. NEWMANN

THE CALL to restructure education sends an ambiguous but serious message. The word *restructure* suggests that the system is so fundamentally flawed that such terms as *innovation* and *reform* aren't robust enough to describe the changes that are needed. But what exactly is the challenge that we face in restructuring? What needs to be changed, in what ways, and for what purposes?

The history of education in America has been punctuated repeatedly with powerful slogans that mobilize the energy of practitioners, researchers, and policy makers. Recent examples include "back to basics," "relevance," "community control," "cultural literacy," and "effective schools." Much of a slogan's appeal rests in its capacity to conjure up multiple meanings that can draw diverse constituencies together in a common cause. While a slogan galvanizes public opinion and focuses energy, thus offering new possibilities for action, its ambiguity brings the great risk that energy will be dissipated in scattered, even contradictory, directions.

Therefore, before committing ourselves to new structures, we should try hard to reach agreement on the nature of the sys-

---

*FRED M. NEWMANN is a professor of curriculum and instruction at the University of Wisconsin, Madison, and director of the Center on Organization and Restructuring of Schools. This article is adapted from a speech presented at the Indiana University Annual Education Conference, held in Bloomington in April 1990, and its writing was supported by Indiana University, by the Office of Educational Research and Improvement of the U.S. Department of Education, and by the Wisconsin Center for Education Research. However, the opinions expressed are the author's own.*

Illustration by [illegible]

tem's failure. For without agreement on the nature of the illness and on the purpose of treatment, the pursuit of radical cures could do more harm than good for students.

Today, reformers tend to address two main issues. The most obvious is the large proportion of students, especially low-income students of color, who fail in school and score poorly on nationally standardized tests. Much of the rhetoric about national education goals reflects concern about this group. Another issue – less popular, but perhaps even more disturbing – is that even those who succeed in school and score well on conventional tests have not been educated to cope successfully with the demands of personal, vocational, and civic life in contemporary society.[1]

Is the point of restructuring to provide a better way of teaching the current curriculum to students who haven't learned it through the old structure? Or is restructuring to be a vehicle to fundamentally change what is taught and how it is taught to all students? These alternative aims for restructuring can reflect conflicting goals of education, and the goals held by those pursuing each aim may lead to the creation of very different types of educational environments. It may not be necessary to choose between these stark alternatives, but the tension must be recognized. It reflects a persistent historical debate about educational ends that has no simple resolution and that even the restructuring movement will not solve once and for all.

Regardless of one's position on the forms of competence that schools should promote and regardless of the fact that large proportions of parents and educators perceive the schools to be working well, there are at least two grounds for seriously entertaining the prospect of fundamental organizational change. First, teachers face the persistent difficulty of engaging students in serious academic work in schools as we know them. Except for a few highly motivated students, most young people complete school only as a ritual. This pervasive disengagement creates massive problems of crowd control for educators and wastes the time of students and staff members alike.

Second, we must recognize the unfortunate record of traditional efforts to teach a virtually limitless list of fragmented pieces of declarative knowledge – e.g., that the issue of slavery divided the U.S. in the 19th century or that squares have equal sides and 90-degree angles. This enterprise, pursued over the ages as the main task of formal schooling, now actually undermines education. When schools try simultaneously to keep up with the explosion of knowledge and to accommodate the host of social demands placed on them, they cannot possibly give students the opportunity to develop in-depth understanding and the capacity for higher-order thinking that are required for success beyond school. The problems of disengagement and of excessive emphasis on coverage are related, and they exact their most tragic toll on students who, because of low income, cultural background, or lack of social support, do not succeed in schools as they are currently organized.

One of the biggest mistakes we make is the order in which we design the components of the education system. First, we set up an organizational structure; next, we plan the curriculum to fit the structure; and finally, we choose criteria for student success. Not until this last step do we articulate in a concrete way the ultimate educational purpose of the system. But by this point the outcomes are largely predetermined by the organizational structures and curriculum. Thus it should not be surprising that schooling seems so dysfunctional.

Moreover, the restructuring movement is about to perpetuate this error, because it, too, is trying to design organizational structures before clarifying purposes and reaching consensus on the educational ends that organizational structures should serve. For example, most of the talk one hears about school-site management, career ladders for teachers, or schools of choice never mentions how these mechanisms will teach students to write about literature, to reason about scientific phenomena, or to learn important geographic facts.

There is no reason to believe that restructuring will produce favorable results for students unless plans for new structures are guided by a vision of educational outcomes that articulates new standards for student achievement and for teaching. In this article I hope to nudge restructuring in this direction. First, I propose *authentic* achievement as the primary goal for students. Then, I advocate the need for *substantive conversation* in teaching, a new standard that is necessary to promote authentic student achievement. Finally, I suggest particular structures for student work that can facilitate authentic achievement and substantive conversation.

**AUTHENTIC ACHIEVEMENT INVOLVES THE CHALLENGE OF PRODUCING, RATHER THAN REPRODUCING, KNOWLEDGE.**

## AUTHENTIC ACHIEVEMENT

The kinds of skills required to earn school credits, good grades, and high scores on typical tests are often considered trivial, meaningless, and contrived – by both students and adults. In contrast, a "restructured" vision of the goals of education seeks to evaluate performance activities that are worthwhile, significant, and meaningful: in short, activities that are authentic.

What criteria help us to recognize authentic forms of academic achievement? Consider the achievements of successful adults – scientists, musicians, business entrepreneurs, politicians, craftspeople, attorneys, novelists, physicians, designers, and so on. What are the characteristics of their work that justify calling their accomplishments authentic rather than contrived and trivial? Can we identify key distinctions between these authentic accomplishments and the work that students complete in school?[2]

People in the diverse fields named above face the primary challenge of producing, rather than reproducing, knowledge. This knowledge is expressed through discourse, through the creation

# *Accomplishments should have intrinsic value beyond their value in assessing knowledge.*

of things, and through performances. We do not expect children to attain levels of competence comparable to those of skilled adults, but we do want students to develop in the same direction.

To progress on the journey toward adult competence, students should set their sights on authentic expressions of knowledge. That is, they must hone their skills through guided practice in producing original conversation and writing, through the repair and building of physical objects, through artistic and musical performance. The conventional curriculum asks students only to identify the discourse, things, and performances that others have produced – e.g., by recognizing the difference between verbs and nouns or between socialism and capitalism; by matching authors with their works; or by correctly labeling rocks and body parts. Certainly, the production of knowledge must be based on an understanding of prior knowledge, but the mere *re*production of that knowledge does not constitute authentic academic achievement.

A second defining feature of authentic academic achievement is its reliance on a particular type of cognitive work that can be described as "disciplined inquiry." Disciplined inquiry, in turn, seems to consist of three features: 1) use of a prior knowledge base, 2) in-depth understanding rather than superficial awareness, and 3) production of knowledge in an integrated (rather than fragmented) form. In highlighting these features, I am not suggesting that students should be expected to make seminal contributions to the academic disciplines, to the professions, or to the arts. However, I am suggesting that students are quite capable of engaging in these forms of cognitive work when they are adapted to the students' levels of development.

1. *Prior knowledge base.* For new knowledge to be significant and valid, it must be based on substantive and procedural knowledge that has been accumulated previously by workers in the field, who establish facts, vocabularies, concepts, theories, algorithms, and conventions for the conduct and expression of inquiry. However, relying on a prior knowledge base does not always mean agreeing with intellectual traditions or continuing to develop them in a linear fashion. The ultimate point of disciplined inquiry is to move beyond existing knowledge – through criticism and the development of new models that are themselves stimulated by prior knowledge. Most of the cognitive work of schools consists of transmitting prior knowledge to students and asking them to reproduce it, rather than helping them to use it to produce new knowledge.

2. *In-depth understanding.* Disciplined inquiry tries to develop in-depth understanding of a problem, rather than only a passing familiarity with or exposure to pieces of knowledge. One masters prior knowledge not primarily to become superficially knowledgeable about a broad range of topics, but in order to develop complex understanding of a smaller number of issues. Such detailed understanding is particularly necessary for the production of new knowledge. In contrast, many of the cognitive tasks of school ask students to show only superficial awareness of a vast number of topics.[3]

3. *Integration.* To produce knowledge, one must assemble and interpret information, formulate ideas, and make critiques that cannot be easily retrieved from the existing knowledge base. All these activities require the ability to organize, synthesize, and integrate information in new ways. Students are unlikely to succeed at such tasks unless they learn to look for, to test, and to create relationships among pieces of knowledge that otherwise appear unconnected.

The third and final distinction between authentic achievements and traditional school achievements is that authentic achievements have aesthetic, utilitarian, or personal value apart from their value in documenting the competence of the learner. Indeed, this distinction may be the most crucial. When people write letters, news articles, insurance claims, or poems; when they speak a foreign language; when they develop blueprints; when they create a painting, compose a piece of music, or build a stereo cabinet, they try to communicate ideas, to produce a product, or to make an impression on others beyond the simple demonstration that they are competent. Achievements of this sort have special value that is missing from tasks contrived only for the purpose of assessing knowledge, such as spelling quizzes, laboratory exercises, or typical final exams. In many cases, the cry for "relevance" is simply a less precise expression of the desire that accomplishments should have intrinsic value beyond their value as indicators of success in school.

To summarize, the idea of authentic achievement requires students to engage in disciplined inquiry to produce knowledge that has value in their lives beyond simply proving their competence in school. Mastery of this sort is unlikely to be demonstrated in familiar testing and grading exercises. Instead, such mastery is more often expressed in the completion of long-term projects that result in the creation of discourse, things, and performances of interest to students, their peers, and the public at large.

Why should we aim toward authentic achievement? I suggest at least two compelling reasons, which respond directly to the problems of the conventional curriculum that I mentioned earlier. First, participation in authentic tasks is more likely to motivate students and to sustain the hard work that learning requires. Because authentic work has value beyond the mere demonstration of competence in school and because it permits more comprehensive use of the mind, students will have a greater stake in authentic achievement. Second, authentic academic challenges are more likely to cultivate the higher-order thinking and problem-solving capacities that are useful both to individuals and to society. The mastery gained in school would then transfer more readily to life beyond school and thus increase the efficiency of our public investment in schooling.

To develop organizational structures consistent with this vision, educators will need to plan "backwards." Beginning with a conception of authentic achievement and of the specific projects that express it, they will need to select appropriate content within the subjects. They will also need to consider the kinds of human interaction — involving teachers, students, and others — that promote disciplined inquiry into the subject. Although I cannot address organizational needs for specific subjects in this article, one aspect of interaction seems critical to authentic achievement in many subjects. This is the need for substantive conversation in instruction, which constitutes a major departure from present practice.

## SUBSTANTIVE CONVERSATION

When the curriculum is geared toward the goal of authentic achievement, new forms of teaching will be necessary. No longer will it be possible for students to succeed in school simply by listening to the teacher and responding to questions that can be answered in half a dozen words. To complete authentic projects, students will need to formulate and to ask questions, to explain themselves to peers and to adults, and to refine their ideas. While many authentic achievements require capacities other than verbal discourse (e.g., working with one's hands, making mathematical estimations, using aesthetic sensitivities), substantive conversation between the learner and some supportive resource — teacher, peer critic, or other knowledgeable authority — is usually necessary. I emphasize substantive conversation, not because it alone is sufficient to develop all forms of authentic achievement, but because it is usually critical for success in authentic tasks and because conversation is the primary medium through which teachers and students communicate.

Substantive conversation differs from conventional classroom talk, in which the teacher usually has only two purposes (aside from disciplinary ones) in speaking to students: to transmit an item of declarative knowledge (usually a definition or a fact) or to determine whether the student can reproduce that item.[4] The student's goal (assuming that he or she desires to cooperate) is to give the answer that the teacher has in mind. The teacher maintains control in that he or she knows, prior to the interaction, what pieces of knowledge must be communicated and what answers will be acceptable.

In a substantive conversation, the purpose of talking is quite different. Each person is trying to express a point of view (often in order to persuade the other), to understand why the other holds a particular viewpoint or interpretation, or to arrive at a solution to a problem that neither has previously solved. This last purpose is the most difficult to fulfill, but it is perhaps the most likely to render the work authentic.

In pursuing any of these ends, the teacher uses talk to integrate subject matter into the student's meaning system, and the student tries to understand how the teacher's messages might empower him or her in the world. In this sense, each is interested in what the other has to say, the course of the conversation is less predictable than in conventional classroom exchanges, and each participant has some sense of ownership. To fulfill these purposes, substantive conversation requires sustained, continuous talk between two or more people. In school as we know it, the two key participants are usually student and teacher, but students can also learn through conversations with their peers, with other adults, and perhaps with well-programmed computers. Teacher/student conversation should remain central. But to expand students' opportunities for expression and feedback, it will be necessary to rely substantially on and to support these other sources as well.

**WHEN THE CURRICULUM IS GEARED TOWARD THE GOAL OF AUTHENTIC ACHIEVEMENT, NEW FORMS OF TEACHING WILL BE NECESSARY.**

What is the connection between substantive conversation and authentic achievement? Begin with the definition: authentic achievement is expressed in discourse, things, and performances that make reasonably complete, integrated statements and that reflect students' production of in-depth knowledge. Then ask the question, How can we instruct students to make such statements? We can show them models of authentic accomplishments of others, and we can give them critical elements of declarative knowledge on which to build. But how will they learn to make their own statements? Substantive conversation is the key. It provides the crucible for practice, for seeking new knowledge that relates to the problem at hand, and for trial, feedback, and revision. In short, substantive conversation forces us to transform declarative knowledge into applied, integrated knowledge.

## STRUCTURES THAT FACILITATE AUTHENTIC ACHIEVEMENT

The Council of Chief State School Officers divides restructuring into four main categories: school governance (including decentralized authority and school choice), reforming the nature and organization of curriculum and instruction, new professional roles for educators, and accountability (especially new methods of assessment and direct state intervention).[5]

This analysis applies most directly to curriculum, instruction, and accountability. But I believe that, unless the entire restructuring agenda focuses more closely on authentic student achievement and on substantive conversation to attain it, restructuring will produce at best only cosmetic changes in students' education.

Assuming a commitment to these goals, what structural conditions seem necessary to achieve them? The conditions I propose below are in no way exhaustive. They outline certain key conditions under which students would work, but they do not provide structures of school governance, detailed systems of assessment, or specific new structures for the professional development of teachers. My argument has implications for these areas as well, and I do support the need for pro-

BEST COPY AVAILABLE

viding teachers with working conditions that enable new forms of professional support. But my central concern here is the structure of the instructional environment.

At least four conditions seem essential for students to complete authentic accomplishments: collaboration, access to tools and resources, worker discretion or opportunity for ownership, and flexible use of time.

*Collaboration.* Achievements outside of school often depend on the opportunity to ask questions of, to receive feedback from, and to count on the help of others, including peers and authorities. In contrast, typical school activities require each student to work alone. Working together is often prohibited, because it is seen as a form of cheating. Of course, it is important for students to learn to work on their own and not become too dependent on others. But if opportunities to cooperate are consistently denied, students will not experience a critical process that adults, both expert and novice, consistently rely on for success.

Useful collaboration depends to a great extent on opportunities for substantive conversation, which itself requires special structural conditions. More time will be needed for teachers to communicate with individual students through sustained talk and writing and for students to talk with one another. Substantive conversation also entails major shifts in the roles of teachers and students. Teachers will function more as mentors and coaches, less as depositories of static knowledge to be reproduced. Students will function more as constructors and producers of knowledge. They will rely on teachers for help, but they will not be mere absorbers or consumers of everything the teacher says. Students will also have to take on the new roles of seeking help from and giving help to one another as they learn.

*Access to tools and resources.* When competent people complete significant accomplishments – e.g., writing books, designing buildings, composing music, or reaching collective bargaining agreements – they usually have made use of a variety of printed materials and such machines as telephones, computers, and laboratory equipment. The information they require is generally too vast and too rapidly updated to commit to memory. In medieval times, before the invention of the printing press, becoming educated required memorizing the contents of key manuscripts, because future access to them was unlikely. But today increasing specialization and the explosion of knowledge create a new challenge. Possession of prior knowledge is critical to disciplined inquiry, but it is now impossible to teach students all the relevant information they will need to complete authentic tasks. The new challenge is to teach students to use tools and resources to find the appropriate knowledge when they need it.

*Discretion and ownership.* Rather than always toiling within the confines of predetermined routines arbitrarily dictated by authorities, the creators of authentic work influence the conception, execution, and evaluation of the work itself. At a minimum this entails exercising some control over the pace and procedures of learning; over opportunities to ask questions and to study topics deemed important; and over constructing and producing knowledge in one's own language, rather than merely reproducing the language of others.

There are, of course, important limits on the extent to which students should control their learning of academic subjects. Certain facts, definitions, concepts, algorithms, and processes of verification must be assimilated according to the standards of the disciplinary fields. But for this kind of learning to be translated into authentic products, students must enjoy some autonomy and discretion in its application.

*Flexible use of time.* The significant achievements of disciplined inquiry often cannot be produced within rigidly specified time periods. Adults working to solve complicated problems, to compose effective discourse, or to design products are rarely forced to work within the rigid time constraints imposed on students, such as the 50-minute class or the two-hour examination.

Standard time schedules stem from bureaucratic procedures designed to manage masses of students and diverse course offerings. Since they have nothing to do with the time requirements of disciplined inquiry, they can reduce the authenticity of student achievement. Achievements in noninstructional tasks, such as journalistic writing, interior design, or medical care, do involve deadlines and time limits. But in such work the schedules tend to be determined more by the nature of the work than by the requirements of institutional management.

The four conditions outlined above

*"I never said 'fusion in my bathtub.' I said, 'I've made a few gins in my bathtub.' "*

**WE MUST RESIST THE TENDENCY TO INSTITUTE ADMINISTRATIVE CHANGES DEVOID OF FRESH EDUCATIONAL VISION.**

suggest at least the following changes in the structuring of students' work.

- To stimulate collaboration and the substantive conversation so crucial to it, more time must be allocated to individualized and small-group instruction.
- Students' access to knowledge must be enhanced by greater use of technology (telephones as well as computers) and by opportunities to learn from sources outside of school.
- Students need to exercise discretion and assume responsibility in the planning, execution, and evaluation of their work, which further underscores the need for individualized and small-group work along with assessment that accepts diverse ways of demonstrating competence.
- Instructional time should be organized to permit more sustained, long-term, and in-depth investigation, in contrast to the fixed time slots designed for survey coverage.
- Assessment tasks calling for the production of discourse, things, and performances will require new procedures and in some cases the participation of individuals who are not members of the school staff to evaluate student performance.

Instituting these changes will contribute to authentic achievement. But they alone cannot guarantee it, for ultimately the quality of student achievement will depend on the quality of conversation that students have with teachers and with one another. For years substantive conversation has been difficult to find in schools, but only partly because of structural constraints. Even if teachers were blessed with more time and much lighter teaching loads, I suspect that substantive conversation would still be rare, because the goals of authentic achievement and substantive conversation also presume major shifts in prevailing conceptions of education. Organizational changes alone will not modify long-standing conceptions of knowledge and deeply socialized habits of didactic teaching that direct most of our educational efforts toward inauthentic forms of mastery.

Just as plans for restructuring should begin with the conditions in which students learn, the next step is to consider the working conditions that teachers need in order to generate substantive conversation. Again, we must not expect organizational structures alone to transform teaching. However, given the complexities of the teaching task, we cannot emphasize enough the need to provide teachers and administrators with sufficient authority, time, and assistance to reflect critically on their practice, to experiment, to fail, and to try again. Beyond enabling teachers to increase the proportion of time they spend interacting with individual students and small groups, we must work to give teachers more opportunity to collaborate, and we must provide for regular, sustained professional development that focuses on the goals of substantive conversation and authentic achievement.

Experiences in restructuring to date reveal three obstacles that dramatize the need for this kind of support. First, teachers, parents, and students who have experienced only the conventional version of education – especially those who have been successful – cling tenaciously to it, even when they have the opportunity to make substantial changes. Faith in conventional practice is fortified by the large proportion of students who do succeed on conventional measures, both in school and beyond. Second, external pressures for district, state, and national accountability, coupled with the need for ways to determine admission to college, continue to sanctify mastery of declarative knowledge as the most important goal of schooling. Finally, the struggle to establish new collaborative roles for teachers is stressful and time consuming; it drains energy and leaves diminished personal resources for thinking about new forms of achievement and curriculum.

The history of education reform suggests that we have yet to see fundamental change on a grand scale.[6] We might argue that the contemporary sense of urgency and the scope of the current restructuring movement combine to offer a unique opportunity. But to avoid the costly mistakes of previous efforts, we must resist the tendency to institute administrative changes devoid of fresh educational vision. If we keep our sights on the targets of authentic student achievement and substantive instructional conversation, restructuring could form a new chapter in educational history – and one well worth the effort.

---

1. Albert Shanker, "The End of the Traditional Model of Schooling – and a Proposal for Using Incentives to Restructure Our Public Schools," *Phi Delta Kappan*, January 1990, pp. 344-57.

2. This complicated matter involving controversial educational values has not yet received extensive scholarly attention. Yet it begs for interdisciplinary analysis of the relationship between formal attempts to educate and the ways that human competence is expressed in noneducational settings. The conception I propose here builds on Harold Berlak, ed., *Toward a New Science of Educctional Testing and Assessment* (Albany: State University of New York Press, forthcoming); Fred M. Newmann and Doug Archbald, *Beyond Standardized Testing: Assessing Authentic Academic Achievement in the Secondary School* (Reston, Va.: National Association of Secondary School Principals, 1988); John Raven, "Toward a Model of Competence, Motivation, and Behavior and a Paradigm for Assessment," in Berlak, op. cit.; Lauren B. Resnick, "Learning In School and Out," *Educational Researcher*, December 1987, pp. 13-20; and Grant Wiggins, "A True Test: Toward More Authentic and Equitable Assessment," *Phi Delta Kappan*, May 1989, pp. 709-13.

3. A commitment to depth over coverage entails no retreat from liberal education into narrow forms of vocationalism, as some might expect. The effort to concentrate on complex understanding welcomes study of a wide variety of disciplines and paths to knowledge. Reducing the emphasis on superficial exposure implies no necessary narrowing of the curriculum as a whole.

4. This analysis is based in part on the work of Martin Nystrand and Adam Gamoran, "Student Engagement: When Recitation Becomes Conversation," in Hersholt Waxman and Herbert Walberg, eds., *Contemporary Research on Teaching* (Berkeley, Calif.: McCutchan, forthcoming).

5. *Success for All in a New Century: A Report by the Council of Chief State School Officers on Restructuring Education* (Washington, D.C.: Council of Chief State School Officers, 1989).

6. Larry Cuban, "Reforming Education, Again, Again, and Again," *Educational Researcher*, January/February 1990, pp. 3-13. K

BEST COPY AVAILABLE

# II-6

# Exhibitions: Facing Outward, Pointing Inward

JOSEPH P. MCDONALD
COALITION OF ESSENTIAL SCHOOLS

Reprinted with permission from the Coalition of Essential Schools, Brown University.

# Exhibitions: Facing Outward, Pointing Inward

It is an old American habit to imagine utopia in the abstract, then search for the instrument — the policy, the remedy, the technology — that might yield it up quickly. But it is just as old an American habit to start more concretely — with images of a Black man voting, of an Appalachian child eating, of a woman firefighter working — then strive, over the long haul, to connect these images and bring them to life.

When it comes to the utopian prospect of creating new schools for the twenty-first century, we in the Coalition of Essential Schools try hard to stick to the second path. Our reason is as pragmatic as it is principled: in school reform, the first path generally offers a fool's journey (Cohen & Ball, 1990; Elmore & McLaughlin, 1988).[1] So, for example, the quick instruments of the 1980s — longer school days and years, more testing of teachers and kids, more course requirements for high schools — seem to have made things worse (Brown, 1991; Timar & Kirp, 1989). Meanwhile, the patient striving of some of the original members of the Coalition of Essential Schools — a creature as well of the 1980s — has turned scanty images of possibility into remarkable schools. These are schools that scout the future — schools like Deborah Meier's Central Park East Secondary School in New York City and Dennis Littky's Thayer High School in Winchester, New Hampshire (Wood, 1992).

Now, at the start of the 1990s, the lure of the quicker path again threatens the capacity of policymakers to focus on the long haul. Thus our schools face a number of proposals that mean to *drive* instruction. These proposals all involve assessment, and sometimes what is called alternative or authentic assessment. Their proponents suggest that clever instruments might be capable of directing schools toward worthy ends — yielding in a stroke the utopia of excellence and equity that many of us seek. Give schools the right test, so the argument goes, and the other elements of new school design will align like metal filaments to a magnet.

Unfortunately, the argument is more attractive than persuasive. It ignores the fact that schools are exceedingly complex organisms, responsive only in perverse ways to outside driving. This is especially true when the driver aims to enhance accountability (Madaus, 1985; Darling-Hammond and Ascher, 1991). Real accountability depends in the end on the resolution of school people to awaken to the effects of regularities that suppress their own best instincts, to come to terms with what Theodore R. Sizer (1990) has called the "essential dailiness of school" — its numbing habit of self-absorption and its preoccupation with simply churning on. One cannot be driven to such mindfulness; one can only be coached into it. Assessment can play an important role in the effort, but only if it aims to point out continually what schools might otherwise overlook to the detriment of their real mission. What is wanted is an assessment system perched powerfully within the school itself and just above the school's other systems — one that from its perch might direct these other systems toward kids rather than toward their own smooth running. This is what the Coalition of Essential Schools means by *exhibitions*.

Ted Sizer (1984, 1992) reached all the way back to the eighteenth century in search of an assessment mechanism that might function in this way. He found at least the possibility of it in a ubiquitous feature of the early American academies and of the common schools that shared their era. The exhibition, as practiced then, was an occasion of public inspection when some substantial portion of a school's constituency might show up to hear students recite, declaim, or otherwise perform.[2] The constituency might thereby satisfy itself that the year's public funds or tuitions had been well spent and that some cohort of young scholars was now ready to move on or out. There is evidence that this satisfaction came cheap. Arthur Powell, who has studied the matter, once told me that the exhibition as practiced in its own time was frequently little more than public entertainment — the equivalent of Friday night football in Odessa, Texas. Wayne Fuller's (1982, p. 212) account captures the flavor:

> Through the years the rural people responded to these exhibitions with enthusiasm. On cold winter evenings, at Christmas time or at the close of the winter term of school, when the snow still blanketed the schoolyard, they crowded into the little schoolhouse and watched their children perform by the light

of the kerosene lamps that cast shadows and an almost eerie orange light about the room. And for the school picnic and exhibition that closed the spring term of school, they brought an abundance of food to the schoolhouse or the grove of trees by the river for the noonday meal and prepared to spend the day talking to their neighbors and watching the exhibition in the afternoon.

For his part, Sizer imagined this old design in something more contemporary than kerosene light; still, he prized its history and aimed to use it to help displace more recent historical influences on the design of the American high school. These are especially the ones that have lent the high school its division of knowledge into subjects, time into Carnegie units, and kids into tracks (Sizer, 1984, 1992; Powell, Farrar, & Cohen, 1985; Hampel, 1986). Features of the exhibition, as Sizer projected them, run counter to these powerful norms: a suggestion that high school kids, like doctoral students, might qualify for graduation on the basis of some *integrative* performance; that they might attempt this performance when they were well enough prepared to attempt it, rather than on some fixed schedule; and that they might *all* be required to do it well, without regard to someone's perception of their abilities, career prospects, or socio-economic status. Together with a number of imaginative school people — like Meier, Littky, and Samuel Billups (of Walbrook High School, Baltimore, Maryland), to name a few of the pioneers of the mid-1980s — and scholars like Arthur Powell (1986) and Grant Wiggins (1988), Sizer launched a theoretical and practical adventure with exhibitions that continues today in the Coalition of Essential Schools.

## EXHIBITIONS AND ACCOUNTABILITY

The exhibition is a device with two related functions — one facing outward and the other pointing inward. It yields accountability in the terms that Linda Darling-Hammond and Carol Ascher (1991) promote: accountability as the product of commitments, policies, and practices designed to press kids toward meaningful achievement and to feed back whatever information the school needs to ensure the success it intends. So the exhibition at its best aims to answer two complexly intertwined questions: What can these kids do now that I've taught them? What can I learn from what they show me?

### *Facing Outward*

There is no more powerful agent of accountability than kids themselves — performing live more or less as the school intended that they should. But this power of live performance depends on having a local audience to assess the rich information it yields, an audience that consists of more than the individual teacher in the privacy of his or her classroom. When kids must show what they know and what they have learned in face-to-face meetings with those who have not taught them but who share an interest in their success, then a certain pitch of accountability can be reached that otherwise is never possible. Of course, accountability systems must transcribe performance into scores or other notation, so that kids' progress may be followed over time and so that others who cannot attend live performances may be informed of kids' standing. But today, most accountability systems provide the transcription without any access to the performance for which it is supposed to stand. One consequence of this is that some among the public come to suspect that the transcription may be groundless.

According to the annual Gallup poll of Americans' attitudes toward schooling (Elam, Rose, & Gallup, 1991), most who are suspicious in this way exempt their own school from the suspicion, and so the politics of accountability tend to collect at the state and national levels rather than at the local one. Yet a closer reading of the most recent poll suggests that this exemption is at best an ambivalent one. It is true that 42 percent of respondents assigned the grade of A or B to their local public schools, while only 21 percent assigned these grades to the public schools nationally. Meanwhile, however, a whopping 81 percent would impose national achievement standards and goals on their local schools, 77 percent would require them to use standardized national tests to assess their community's kids, and 68 percent would have their local schools follow a national curriculum. If there is so much local satisfaction, how come these other figures? My answer is that most local communities trust their local educators — urban communities notably excepted — but nonetheless yearn for reassurance in such uncertain times. In the Coalition of Essential Schools, we think this reassurance may best be supplied by means of public access to live performance. Exhibitions can provide such access.[3] When they do, we say that they face outward.

Of course, the politics of accountability (including the demands for national testing, national curriculum, and so on) are fueled by more than local concerns. An important factor too is the (well-founded) suspicion on other levels that many local schools and, indeed, many local communities have at best an inchoate sense of twenty-first-century standards. Some schools and communities clearly still care more that their kids spell well or memorize the list of state capitals than that they acquire intellectually powerful habits of mind. This fact — clearly not in the *national* interest — emboldens some critics, scholars, and policymakers to suggest that real local needs — as contrasted with perceived ones — may indeed be better attended to on some "higher" than local level.

In the Coalition of Essential Schools, however, we object to this conclusion on both ethical and political grounds. To us, it seems right to begin the work of ratcheting up expectations on the local level, since expectations are invariably entangled with values, and values legitimately vary somewhat from one American community to another. It may also be politically prudent, since these values often vary in explosively political ways. Moreover, if we are right about the substantial role played in the politics of accountability by local yearning for it — if, indeed, people *do* want to see their kids' achievement for themselves, despite their apparent satisfaction with local schools — then they will be no more satisfied by individual profiles from a redesigned National Assessment of Educational Progress (NAEP) or by criterion-based scores on a voluntary national exam or state-scored portfolio than they are satisfied now by report cards and Lake Wobegon statistics.

This is not to deny that parents also want to know how their kids stack up against other kids within the school, in the next town or state, even in other countries. Nor is it to argue against new roles for NAEP or against other national assessment efforts or state monitoring and moderating. Sensible accountability policy for the 1990s and beyond may well require imaginative expansion in each of these areas, and it must come to terms with what I would argue are the extraordinarily comparative concerns of this generation of parents.[4] On the other hand, attention to these things can easily diminish accountability rather than enhance it. This will happen especially if attention to individual student assessment displaces all attention to the quality of schools' efforts to press for achievement and if local schools are to be perceived as the mere collectors of accountability data rather than as the heart of the accountability system (Darling-Hammond and Ascher, 1991).

By contrast, in the Coalition of Essential Schools we imagine a nationwide accountability that is rooted in local efforts to press hard and to take stock, supported and monitored by district and state, fed by perspectives on achievement and equity that are developed and refined on the national level. Such a system would require of each school that it open itself to inspection, report annually to its community like a corporation to its stockholders, disclose what all its indicators of achievement reveal. These indicators should include not only common measures like attendance and drop-out rates and comparative achievement data, but uncommon ones like longitudinal indicators of achievement and the living evidence of the school's current candidates for graduation showing their stuff in exhibitions and standing for public examination.

### *Pointing Inward*

Besides helping schools face outward, exhibitions at their best also point inward. Darling-Hammond and Ascher (1991) assert that accountability depends ultimately not only on a school's responsiveness to its constituencies but also on its capacity to generate and act upon information regarding its own inner workings. Some schools within the Coalition of Essential Schools have recently discovered the power of exhibitions in this respect. Prominent among these are the schools that approached restructuring from the "end" rather than the "front."

While many member schools have tied their innovations to a pioneer cohort moving ahead grade by grade or to a pilot project designed to get the kinks out of a structural innovation and then expand it, "end-first" schools have often dared to establish exhibition systems right away. These systems vary in design from school to school — creatures of the institutional forces, histories, and personalities that have built them. But in each case, they offer the school more than a showcase for outcomes. They also offer a perspective, previously unavailable, on the systems that produced these outcomes of an innovation. And they offer a means for the school to take stock. One can think of them as a kind of *platform* — one that serves as a stage for student performance and, in another respect, as a school's lookout on its own systems. By means of such platforms, schools may see the real effects of their current work and also adjust the work's dynamics so as to ensure better effects in the future. The Coalition calls this process "planning backwards," which is to say, orienting a school by the compass of its graduates' performance (McDonald, 1991).

## TWO PLATFORMS FOR PLANNING BACKWARDS

One member of the Coalition, Sullivan High School in Chicago, requires that all its seniors (152 last year) demonstrate their ability to engage thoughtfully with difficult texts by participating successfully in a ninety-minute seminar on several such texts, then writing an acceptable essay on the ideas embedded in the texts. This is the platform. Its construction owes much to the influence on the school of Mortimer Adler's (1982) *Paideia Proposal*.[5] It has a simple design, but one that has had more than an additive impact on the rest of the school's structures. For one thing, the whole faculty is involved in leading the seminars and scoring the seniors' participation, in coaching the seniors during the two weeks they have to grapple with such texts as Freud's "Why War?" and Thucydides' "Melian Dialogue," and in scoring the essays. Thus, the whole faculty feels pressure to understand the scoring rubrics and the pedagogy that underlies them and to practice this pedagogy in their classes regardless of grade and subject.

Initially, this pedagogy was widely practiced only in the school's Paideia Impact Program — a magnet program within the larger school. In insisting on the establishment of the schoolwide seminar-exhibition, Principal Robert Brazil in effect mandated the extension of this program schoolwide. He entrusted the implementation of the mandate, however, to an influential group of teachers. The consequence was that the mandate stimulated rather than quashed professional esteem. So far, the effort has led to peer-training mechanisms in seminar-based and writing instruction, to scheduling changes that allow weekly double-blocks for seminars and team-teaching, to new forums for school governance, to more experimentation with coaching as a method of teaching, and to design efforts aimed at requiring a senior project in math and science to complement the senior seminar.

Another school in the Coalition, Boston's English High School, designed a different platform, the product of what might be called the principal's epiphany. Sid Smith found himself one graduation day handing out diplomas to several kids who he knew could not adequately read and write, but who had nonetheless completed graduation requirements. The next year, he and the faculty instituted a new graduation requirement: in addition to earning a certain number of course credits, every senior must write a "position paper" that demonstrates a capacity to think through a complex issue, back up an opinion with reasoning and reference, and do it in writing that is persuasive and free of error.

From this simple platform English High has learned to see itself more clearly and to wonder productively about what it sees. Why is it that some kids formerly graduated without being able to do what the position paper demands? How often in ninth-grade courses, for example, are kids required to take positions and defend them orally and in writing? How much classroom time is instead dominated by teachers' talking and kids' one-word responses to their questions?

This seeing and wondering have led to a number of changes, present and prospective: peer-training mechanisms again, greater collaboration across subjects and grades, introduction of coaching mechanisms to supplement direct and whole-class instruction, and exploration of emerging technology to provide students richer and more independent access to information. Another result has been some elaboration of the original platform: the addition of a "defense" to the paper requirement, the addition of earlier paper requirements at each grade level, attempts to introduce quantitative thinking into the papers and the defenses, a design effort aimed at supplementing the paper requirement with various project requirements.

It is important to note that exhibitions worked no magic in either of these cases: these platforms were designed to be instruments of leadership, not salvation. It is also important to note that this leadership originated with the principals in both cases but soon became collective.

## WHAT EXHIBITIONS POINT TO

Exhibitions that sit on such good platforms as the two described above tend to point to or highlight five different dimensions of school life, each of which constitutes a crucial arena for school reform.

### *Results and Goals*

First and fundamentally, exhibitions like the two above point to results, the live results of kids performing — a record more startling in its impact than letters on a transcript or deviations from the norm of a standardized achievement test. The tonic is bracing for an enterprise used to indecipherable effects, one tuned more to process than product. Inattentive to live results, schools adjust easily to the grip of their own running on, where what matters most is to keep moving, where achievement is measured in time served. Confronted by results, schools may reflect on the efficacy of the systems that produce these results.

Next, and also fundamentally, exhibitions point to goals. This is also a bracing tonic. Although talk of goals typically fills many hours of the professional education of teachers, goals in practice are commonly secret and indistinct, as Powell (1986, p. 64) suggests:[6]

> A certain vagueness about desired outcomes has been found useful by teachers for many different reasons. Individual teachers often have their own goals. Working on them privately is one dimension of professional independence. Working on them collaboratively is often regarded as unwanted intrusiveness. Teachers, in addition, often resist too much "outcome specification" because of its reductionist implications. They argue that, in general education as distinct from skill training, the *unpredictability* of many outcomes is inevitable and desirable. Moreover, some fear that too much specification might lead either to unacceptable failure rates or unacceptably low minimum standards. Vagueness and individual autonomy regarding assessment prevent these undesirable results. (emphasis in original)

In the absence of professionally articulated goals, others slip in. These typically apply differentially to kids. The most salient for some may be getting a "4" on the AP English exam and an admission letter to Stanford, while those for peers down the hall may be getting "through" high school with minimal reading skills and no more pregnancy. Because exhibitions by design involve all the school's candidates for graduation and occupy a sphere of school life apart from — though related to — patterns of instruction, they have a power little evident elsewhere in high school to point out the intellectual shallowness of such aims as these, as well as the severe discrepancy between them. They may prompt a school to ask itself what really matters — what *all* kids should know and be able to do to live well in the American twenty-first century. One benefit of asking such a good question is that it may well direct the questioner to the counsel of interested outsiders. In the best of circumstances, these will include not only parents, partisans of local culture, and the local business community, but also those more remotely outside — scholars in the disciplines, others with articulate views of the economic, cultural, political, and moral national interest.

### *Professional Development*

In pointing to results and goals, exhibitions implicitly alert us to the prevalence of the "fudge factor" in schooling, the tendency *not* to worry about results and goals. The fudge factor is a predictable effect of the uncertainties that fill a teaching life and of a chronic habit within the discourse about schooling to discount these uncertainties (Cohen, 1988; McDonald, 1992). Many speak about teaching as if teachers could and should work like technicians — diagnosing problems, applying solutions. But this way of working is really impossible in teaching, where so much depends upon the volatility of interpersonal relations and the complexity of human understanding.

On the other hand, setting schoolwide goals, taking collective responsibility to reach them, and honestly facing results are powerful means for coping with uncertainty. But they are aspects of craft rather than of technology. They invite teachers to value teaching as something more than the application of other people's inventions — indeed, to appreciate it as the workshop of their own inventions. In an important sense, exhibitions may be said, therefore, to point to the responsibility of every school to pursue the professional development of its faculty within the fabric of school life rather than outside it. An exhibition platform may be said to work well to the extent that its features provide teachers with opportunities, for example, to collaborate with colleagues as well as outsiders in the design of prompts and scoring rubrics and other materials, to collaborate also in the establishment and maintenance of standards, and to tinker as needed with the operating systems of the school. As every teacher knows who has ever enjoyed such opportunities at work, these constitute the best professional development available.

### *Standards and New Ways of Teaching*

Schools that dare to set up platforms for exhibitions commit themselves to prepare all their students to perform upon these platforms in some manner that exceeds the ordinarily minimalist focus of the high school — where persistence alone counts most heavily. In fact, people associated with the Coalition of Essential Schools often speak of exhibitions as "demonstrations of mastery." But what is *mastery*, after all? When is a particular performance good enough, and when is it not? And how can a school reliably equate "good enough" across the performances of many kids engaging many different tasks over the course of many years? These are questions of standards, and exhibitions invariably point them out.

Schools that wrestle with standards typically begin by building internal mechanisms for teachers to talk with each other about the quality of student work. This is a crucial first step, though it is not enough by itself. The maintenance of high standards in school also depends on teachers' access and openness to outside expertise. Because many exhibition platforms insist that the student-exhibitors forage for information and insight beyond school walls, the faculty that judges their work must be willing also to distrust the sufficiency of inside resources — including the resources available from their own previous learning. So exhibitions may be said to point to a different conception of teaching than that represented in the prevalent metaphor of teaching as the delivery-of-instruction.

People in the Coalition often use the alternative metaphor of teaching as *coaching*, but the word may be too familiar to do justice to the sense I mean. Teachers who prepare kids to write position papers, for example, and who support them in the writing, must be themselves prepared to function as orienteers to territories of knowledge they know little or nothing about, as advisors to students whose chief intellectual mentors on particular projects are outsiders the teacher does not even know, as experts on the *process* of knowing rather than on its substance. Teachers who prepare students to undertake senior projects of significant scope and depth must learn to teach more frequently in ways that give even their youngest students practice in staking out and managing their own learning initiatives and in planning and performing whole routines. Trained and socialized in the pedagogy of the dole, teachers can have a hard time figuring out how to teach under these new circumstances — how to let go enough while still providing safe boundaries, how to help without directing, how to distinguish a timely intervention from a controlling one, and so on.

### *Nature of Knowledge and Learning*

Exhibitions also tend to point to new conceptions of what knowledge is and of how minds may grasp it. Because exhibitions are integrative acts, often calling upon a candidate for graduation to synthesize what he or she has learned in more than one domain and put it to use, they highlight a view of knowledge as integrated perception, cognition, and use. This view happens to coincide with that of contemporary learning theorists as to how knowledge may best be acquired; that is, by active contact with it in context and use. Resnick (1987), for example, argues that the legacy of early twentieth-century psychology in decomposing and decontextualizing knowledge for better delivery has handicapped teachers and learners alike.

The challenge of undoing this legacy, however, is daunting. At issue is not just the organization of teaching into rational surveys and stepwise experience, but also the division of the curriculum into mutually exclusive subjects and the tendency of schools to ignore habits and skills that transcend this division. At their best, exhibitions point all this out. To address the question "How pure is Baltimore's water?" (as one exhibitor at Walbrook High School recently did), one may reasonably explore chemistry, engineering, biology, geology, politics, public health, public policy in general, and even history, literature, philosophy, and the arts. And to address the question well, one will have to know how to do some things that are the province of none of these subjects alone — for example, how to use a question like a searchlight and like a knife, how to keep track of emerging insight, how to find out what you need from people and books that go way over your head, how to work your way out of intellectual and emotional cul-de-sacs, and so on.

### *Incentives and Means*

Having managed to pull together a vision of what it wishes its students to become, the school that dares to plan backwards from exhibitions must also consider how it can help these students make that vision their own. In effect, the exhibition is a kind of compact by which school and student agree to work toward this vision. It points the way. The way can hardly be found, however, unless the student shares the vision, sees what is pointed out. So the school must confront some questions that most schools never confront: How should the school portray for kids what it hopes they will become? How can they ensure that such images will be compelling and empowering rather than threatening? How much student voice will the school admit to the process of inventing and discussing these images? How can the school ensure that the images honor rather than stigmatize difference? What benchmarks for growth should the school provide, and how can they be provided so that kids will take them as encouraging rather than discouraging?

Such questions arise outside the ordinary conception of school as a processing machine. They suggest an alternative conception in which schools and kids collaborate on the terms of their engagement with each other. They also provoke healthy skepticism about the traditional means of instruction. If what is wanted is kids who think things through, write well, speak persuasively, act with confidence, work well with others, trust their intuitions, and so on, how much practice with these things can they gain by sitting silently in large groups most of their days? So exhibitions point to other means — projects, advisory groups, coaching groups, seminars, studios, workshops, school governance mechanisms, mentorships and apprenticeships outside the walls of the school, and so on.

It just so happens that most of these other means implicitly require that the school invent better custodial arrangements than the one prevalent now — the one that requires that kids sit continuously in small desks with a teacher's eye on them. Thus, exhibitions ultimately point schools toward a reconsideration of what Seymour Sarason (1982) calls the "constitution" of schooling. This is perhaps the most important pointing, the one that may, if we're lucky, give us genuine school reform in the 1990s rather than some other false image of it. As Sarason (1982, 1990) suggests, our persistent failure to restructure the power relations of schooling foretells our failure to achieve any other kind of meaningful restructuring. What we need are schools that are genuine communities where the authority to maintain norms of personal and intellectual respect is diffusely shared and where habits of decency run deep.

## REFERENCES

Adler, M. (1982). *The Paideia Proposal.* New York: Macmillan.

Brazil, R. D. (1988). *The Engineering of the Paideia Proposal: The First Year, 1984-1985.* Champaign, IL: School Design Group.

Brown, R. G. (1991). *Schools of Thought.* San Francisco: Jossey-Bass.

Cohen, D. K. (1988). "Teaching Practice: Plus Que ça Change." In *Contributing to Educational Change,* P. W. Jackson (Ed.), pp. 27–84. Berkeley, CA: McCutchan.

Cohen, D. K. & Ball, D. L. (1990). "Policy and practice: An overview," *Educational Evaluation and Policy Analysis 12,* 233-239.

Darling-Hammond, L. & Ascher, C. (1991). *Creating Accountability in Big City School Systems.* New York: Teachers College, Columbia; ERIC Clearinghouse on Urban Education.

Elam, S. M., Rose, L. C., & Gallup, A. M. (1991). "The 23rd Annual Gallup Poll of the Public's Attitudes toward the Public Schools," *Phi Delta Kappan 73* (1), 41–56.

Elmore, R. F. & McLaughlin, M. W. (1988). *Steady Work: Policy, Practice, and the Reform of American Education.* Santa Monica, CA: Rand.

Fuller, Wayne E. (1982). *The Old Country School: The Story of Rural Education in the Middle West.* Chicago: University of Chicago Press.

Hampel, R. (1986). *The Last Little Citadel.* Boston: Houghton Mifflin.

Kraushoar, O. F. (1976). *Private Schools: From the Puritans to the Present.* Bloomington, IN: Phi Delta Kappa Educational Foundation.

Madaus, G. F. (1985). "Test Scores as Administrative Mechanisms in Educational Policy," *Phi Delta Kappan 66*, 611–617.

McDonald, J. P. (1991). "Dilemmas of Planning Backwards," Studies in Exhibitions, No. 3. Coalition of Essential Schools, Brown University, Providence, RI. (to appear in *Teachers College Record*).

McDonald, J. P. (1992). *Teaching: Making Sense of an Uncertain Craft.* New York: Teachers College Press.

Powell, A. (1986). "Exhibitions of Mastery: Some Preliminary Considerations." Research paper prepared for the Coalition of Essential Schools, Brown University, Providence RI.

Powell, A., Farrar, E. & Cohen, D. (1985). *The Shopping Mall High School.* Boston: Houghton Mifflin.

Resnick, L. B. (1987). "Learning In School and Out," *Educational Researcher 16* (9), 13–20.

Sarason, S. (1982). *The Culture of School and the Problem of Change*, 2nd ed. Boston: Allyn & Bacon.

Sarason, S. (1990). *The Predictable Failure of Educational Reform.* San Francisco: Jossey-Bass.

Sizer, T. (1964). *The Age of the Academies.* New York: Teachers College Press.

Sizer, T. (1990). "Assessment: A Cautionary Meditation from the Schools." Paper presented at the annual meeting of the American Educational Research Association, Boston.

Sizer, T. (1984). *Horace's Compromise: The Dilemma of the American High School.* Boston: Houghton Mifflin.

Sizer, T. (1992). *Horace's School: Redesigning the American High School.* Boston: Houghton Mifflin.

Timar, T. & Kirp, D. (1989). "Education Reform in the 1980's: Lessons from the States." *Phi Delta Kappan 70* (6), 505–511.

Wiggins, G. (1988). "The 'Exhibitions' Project." Research paper prepared for the Coalition of Essential Schools, Brown University, Providence RI.

Wood, G. (1992). *Schools That Work.* New York: Dutton.

## NOTES

[1] One may distinguish these paths by saying that the first is for those who think the end of the revolution is near at hand, while the second is for those who want to start a revolution. Cohen (1988, pp. 53–54) suggests how such a difference in perspective may account for the difference in approach:

> The early work calls for exploration of alternatives, invention of many forms knowing that only a few may succeed, experimentation, and creation of examples that suggest the possibilities of change. . . . But at the end of a long revolution there is little time for exploration, experimentation, and invention. The top priorities are to take possession of disputed institutions, to consolidate power and ideology, and to dispose of old enemies.

[2] The heyday of the academies and their exhibitions was the mid-nineteenth century, just before the rise of the high school. Thereafter, many disappeared, though some were transformed into colleges (Mt. Holyoke, Washington and Lee) or normal schools, and some lived on (the Phillips Academies, Emma Willard School). The academies overall had some of the characteristics current radical choice advocates desire: a mingling of private and public funding, great varieties of aim and form, grassroots responsiveness to the "market" of their time, and a diverse clientele. On the other hand, they also had some of the characteristics that opponents of radical choice plans fear: transience of all sorts, an abundance of poorly qualified teachers, and highly uneven programs (Sizer, 1964; Kraushoar, 1976).

[3] There are certainly other mechanisms that might also provide the needed access. One, for example, might involve an American variation on Her Majesty's Inspectorate, the 150-year-old British institution that renders independent and thoughtful accounts of life in schools for the benefit of policymakers and the public.

[4] My argument is that this parental generation, the baby-boom bulge, is projecting its own experience of intense competition in all spheres of life — schooling, employment opportunity, real estate, etc. — onto a generation that is demographically

quite different and unlikely to experience anything like the competition their parents face.

[5] Sullivan's principal, Robert Brazil, who is director of the Paedeia Institute of Chicago, has written a book that explores the application of Adler's concepts to a city high school. See Brazil (1988).

[6] This section is indebted to Powell's essay, which, as he put it, mapped the territory of exhibitions. My sense of that territory, six years later, is only a little different from his.

# III

# *Developing Student Learning Goals for the 21st Century*

# III-1

# Developing Student Learning Goals for the 21st Century

## Introduction

Robert E. Blum
Northwest Regional Educational Laboratory

### New Goals for a New Age

Over the past decade, educators and parents alike have increasingly recognized the need to improve the school performance of all students, and concern about both the equity and quality of results continues to grow. The evidence cited to show progress—or lack of progress—has included standardized test scores and social behavior factors such as attendance and compliance with school rules. The percentage of students achieving above the national norm is a common indicator used to show improvement or lack thereof in student academic achievement. Indicators such as the percentage of students attending school each day and the number of referrals for disciplinary purposes have been used to show improvement in student social behavior. On measures such as these, some schools and districts have demonstrated improvement.

By looking to these commonly accepted measures of student achievement and behavior, we assume that what they measure is important for students' future success. More and more people, however, believe that current measures include only a small part of what students should learn for success in the 21st century. The goals for student learning are changing. At all levels of education, people are reconsidering and talking about what students should learn in school and how well they should learn it. What do students need to learn to assure their future success and the continued success of the United States as a world leader?

### Perspectives on Goals for Student Learning

There are many perspectives on what students should learn in school. One strong theme is that students should be prepared for productivity in the workplace. This point of view is presented by **Anthony P. Carnevale, Leila J. Gainer, and Ann S. Meltzer** in a U.S. Department of Labor report summary immediately following this introduction.

Yet another view is that students must develop into active participants in our democratic process. The late **Ernest L. Boyer** argues that complacency about political processes is as great a threat to our society as lack of preparation for productivity in the workplace. He describes both the need for civic education—newly defined—and the attitudes and skills young people need to participate fully in our democratic process.

Bringing meaning to one's life should be a primary aim of schools, according to **Elliot W. Eisner**. Valuing learning and pursuing learning outside the school, formulating questions and problems and determining strategies to solve them, developing multiple forms of literacy, wondering and imagining and the creativity that results from such use of time, establishing community and the place of the individual within it, and helping each person develop a "personal signature" are the important goals of schooling. Eisner points out that it is not scores on standardized achievement tests that make a difference, but rather the kinds of critical skills that young people can use in tasks outside the classroom and throughout their lives.

Violence and crime are on the minds of people throughout the nation. Citizens do not feel safe, and neither do people from other countries who come to the United States for business or pleasure. **Alfie Kohn** argues that it is not enough for the schools to produce good learners, they must also produce good people. Developing in students the qualities of generosity and a commitment to the welfare of others is an important responsibility of the schools. A key point made by Kohn is that teaching a unit on values or moral reasoning is not sufficient; rather, the whole school culture must model prosocial behavior demonstrating caring, supportive relationships among all the people in the school community and beyond.

A strong voice throughout the nation is asking schools to return to the basics, meaning "readin', (w)ritin', and (a)rithmetic." **Paul Berman and Driek Zirinsky**, however, report that a statewide investigation in Idaho revealed widespread support for "new basics" or "core literacies," including reading, writing, mathematics, sciences and computing, communication, problem solving, assessing and using information and global awareness. Their report goes on to say that students who want to be successful in the next century must also develop a "can-do" attitude, collaborative ability, and adaptability.

The contemporary world is experiencing an information explosion. Knowledge is growing exponentially. Schools cannot keep up with the facts, and technology is making it unnecessary for us to memorize them. This explosion of easily accessible information has led to an increasingly widespread belief that learning to learn is the fundamental purpose of schooling. In their paper, **Beau Fly Jones and Todd Fennimore** weave together research reflecting philosophical, cognitive, and multicultural perspectives to describe successful learners as knowledgeable, self-determined, strategic, and empathetic individuals. Successful learners amass substantial content knowledge, think with fluency, expand their tools to engage in meaningful learning, have a repertoire of effective strategies for pursuing their own learning, and are able to communicate with others and view themselves and the world from the perspectives of others.

The views outlined above are not comprehensive, and people have not yet agreed on what students should learn in school. In reading across the perspectives, though, one does develop a sense that schooling needs to encompass more than the traditional subjects. But schools cannot do everything. With guidance from national and state goals, each community must determine what is important for its children to learn to be successful in the future.

## Statement and Definition of Student Learning Goals

Everyone is developing goals for students' learning. We have the national goals set by the state governors and the president through the Goals 2000 process. Not to be outdone, many states have developed or are in the process of developing goals for student learning. School districts and schools are developing goals as well. Statements of student learning goals establish the mission of the schooling enterprise. The work going on at all levels is redefining the intent of schooling from a tight focus on a one subject to a broad set of skills and knowledge needed for success in the future. Schooling is becoming less concerned with developing the knowledge to move ahead in school and more concerned with learning how to access knowledge for use in decision making and problem solving related to life outside school. Because education is increasingly important for success in adult life, schools are being asked to link what happens in school to our work and life outside the schoolhouse.

Three examples in this section illustrate student learning goals developed by local school districts. The first is from the **Snoqulamie Valley School District** in Washington. The mission statement and goals show the mixing of student learning goals with organization goals. All the goals are important to the "stakeholders" in Snoqualmie Valley, but not all of them are focused directly on what students should learn. The second direction statement is from the **Salem-Keizer School District** in Oregon. This example includes a clear, broad statement of student learning goals and statements describing the partnerships and organizational effectiveness necessary for all students to be successful learners. The Salem-Keizer direction statement also shows the relationship between local and state goals. The third example is from the **Aurora Public Schools** in Colorado. The Aurora goals show a strong focus on adult life roles with content knowledge being woven into the goals.

In early 1989, the **Commission on Maine's Common Core of Learning** began a journey to determine the knowledge, skills, and attitudes that Maine students need by the time they graduate from high school. Sections I and II of the Commission's report describe the process used to develop the common core, the reasons for changing schooling in Maine, and the goals to be achieved through the schooling process. This example of state-level goals is thorough and illustrates the blending of subject area goals with applications beyond high school. The four major goal areas—Personal and Global Stewardship, Communication, Reasoning and Problem Solving, and The Human Record—cut across subjects and provide the basis for integrating knowledge and learning.

The federal government has established in law eight educational goals for the country. The initial six goals were established in 1989 through a consensus process involving governors and federal officials. In 1993, the list was expanded to the current eight, including school readiness, school completion, student achievement and citizenship, teacher education and professional development, mathematics and science, adult literacy and lifelong learning, safe, disciplined, and alcohol- and drug-free schools, and parental participation. The materials in this section on the **National Goals** provide the full goal statements and objectives for each of the goals. As will become apparent from a review of the goals, the National Goals are not all statements of student performance. Goal 1, School Readiness, for example, describes resources to be made available to young children and their parents and does not list or describe the skills and knowledge children need to be successful in school. Goal 3, Student Achievement and Citizenship, on the other hand, describes expected student performances in a broad range of subjects. The National Goals contain both student performance goals and system performance goals.

The broad, general statements of student learning goals are the starting point, but the statements need definition. What does it really mean to prepare young people for active participation in our political process? What does it take to prepare young people to be productive workers? Expanded definitions provide rich descriptions of the qualities, characteristics, skills, and knowledge that students should develop through schooling. Expanded definitions include more than knowledge and skill—the attributes we have typically thought of as developing in school. Integral to most lists of student learning goals is the conviction that it will take more than traditional skills and knowledge to be successful in the 21st century. For example, a "can-do" attitude is included in the Idaho list above.

Expanded definitions take many forms. Some are lists and others are narrative descriptions. The most common approach to defining student learning goals has been to break each goal into several levels of increasingly more detailed parts. The "reductionist" approach may have served schools well in the past, but our rapidly expanding knowledge base makes it more and more difficult to focus our attention on detailed levels of skills and knowledge as we bring meaning to what students should learn in school. The trend now is toward a more holistic approach to defining important student learnings. As Margaret Wheatley describes in her 1992 book *Leadership and the New Science: Learning about Organizations from an Orderly Universe*:

> Scientists in many different disciplines are questioning whether we can adequately explain how the world works by using the machine imagery created in the seventeenth century, most notably by Sir Isaac Newton. In the machine model, one must understand parts. Things can be taken apart, dissected literally or representationally (as we have done with business functions and academic disciplines), and then put back together without any significant loss. The assumption is that by comprehending the workings of each piece, the whole can be understood. The Newtonian model of the world is characterized by materialism and reductionism—a focus on things rather than relationships and a search, in physics, for the basic building blocks of matter.
>
> In new science, the underlying currents are a movement toward holism, toward understanding the system as a system and giving primary value to the relationships that exist among seemingly discrete parts (pp. 8-9).

If we apply the concepts presented by Wheatley to the development of the student learning goal of writing well, we can probably conclude that developing skill in one dimension of writing—voice, for example—does not lead to high-quality writing. Neither will the development of high-quality writing skills assure that a person can communicate well with others every day. There seem to be an essence of good communication that spans the various types of communication and is not present in the detailed skills contained in any one type of communication. As our word becomes increasingly more complex, working to find the essence of what is important to learn, rather than proliferating details by attempting to reduce what students should learn to the lowest common denominator, is critical.

Oregon, like several other states, has developed student learning goals, referred to as Certificate of Initial Mastery Outcomes. The material on **Oregon Certificate of Initial Mastery Outcomes** describes the goals in some detail, including narrative description and a list of key factors or traits. The **Extended Outcome Definitions** provide a basis for developing scoring rubrics for each goal area.

Some descriptions connect student learning goals with the curriculum. **The Carnegie Foundation** report is one such description. Goals such as critical thinking, healthful lifestyles, and active citizenship are connected with approaches to learning like studying interdisciplinary themes, role playing, and service learning. Descriptions such as this one provide insights about what students should learn and methods that have the potential to foster the desired learning.

## Standards for Student Learning

"Standards" is becoming one of the most commonly used terms in the language of school improvement and restructuring. To national and state legislators and policymakers, the term standards is central to their thinking and talking about change in schools. The term is used less frequently by people at the local level—teachers, school administrators, parents, and others. Although people do not agree on the nature of standards, they do agree on the basic meaning of standards. According to Webster's dictionary, a standard provides the basis for comparison or judgment. The *National Science Education Standards* indicate that standards are statements that can be used to judge quality. And the *Curriculum and Evaluation Standards for School Mathematics* say that "a standard is a statement that can be used to judge the quality of a mathematics curriculum or methods of evaluation."

Opinions vary on the types of standards and how they should be used. The **National Science Education Standards Discussion Summary** describes five types of standards related to science education, including program standards, teaching standards, content standards, assessment standards, and system standards. None of the types of standards relates to the quality of student performance in science. All the various types of standards are important in working toward quality schooling for all children, youth, and young adults, but standards for student learning in important goal areas are essential.

Student performance standards provide the basis for judging the quality of learning attained by each student. If we assume, for example, that a standardized test of reading measures important student learnings, then the quality of student learning can be determined in part by looking at test results. In this example, a student who scores in the top quartile on the test would have achieved higher quality learning than a student who scores in the bottom quartile on the same test. In another, perhaps more authentic, example, a student who scores 4 on a scale of 4-high to 1-low on each of seven traits in a writing performance test would have demonstrated exceptional quality in writing. On the other hand, a student who scores all 1's on the same writing performance test would have demonstrated very low performance. Continuing the example of student writing performance, acceptable performance standards can be established. A school may decide that all students should score no lower than 3 in all 7 traits assessed in the writing performance test.

Although student learning is the "bottom line" for many people, using student performance information to judge the quality of school performance can be problematic. **Linda Darling-Hammond** explains some of the issues related to using student performance information to judge school quality. One of Darling-Hammond's major points is that a good accountability system includes other types of standards. Darling-Hammond's essay leaves the impression that using student performance standards is so problematic that we should rely most heavily on standards of practice. This position leads away from a focus on what and how well students are learning, which I believe is counterproductive to a mission of quality and equality in learning.

## Who Decides and By What Criteria?

Stakeholders at the local level are accountable for the learning success of all students in the community, and they must decide what is important for students to learn in school. States have the legal responsibility for the schooling of our children, youth, and young adults. Local school districts are the legally constituted entities charged by states with the responsibility of providing schooling for all children, youth, and young adults in the communities they serve. Schools are where the action is, so it is the local stakeholders who must decide what students learn.

The first criterion for deciding what students should learn is that the learning goals themselves must be *locally developed*. School districts must bring their communities together and lead people through processes that result in communitywide consensus on what is important for students to learn. Districts must assure that all key stakeholders in the district and in each school are involved so that commitment to the goals remains high even after the development process.

This community-based effort might initially seem to be a "loose" process that could result in "lowest common denominator" goals. Thus, the second criterion is that goals must be *soundly based*. This criterion gives the process strength and quality. It means that the stakeholders who are deciding what is important for students to learn must use a broad base of research and theory to support the identified student learning goals. Through a learn-design process, groups study the national goals, state goals, various perspectives on what is important for students to learn, descriptions of alternative futures in which our youth will live, societal trends, and other key information, as they bring their own perspectives into the consensus development process. The process involves common learning by all participants, reflection on the learnings, sharing of personal and group perspectives, reflection to detect commonalities, drafting of goal statements, review and revision of several drafts by increasing numbers of people, and, finally, formalizing statements of student learning goals.

Again, districts lead the process, and all stakeholder groups from all schools in the district are engaged in setting the goals. Using the criteria of local development and a sound basis for the goals will assure stakeholder consensus and commitment.

## Processes for Identifying Goals and Standards

Thomas A. Olson lays out a comprehensive approach to developing student learning goals and a district mission. The process involves relatively large numbers of people from all "key stakeholder" groups—students, parents, teachers, administrators, support staff, community members, and business representatives—in a two-stage process. Stage one is concerned with developing student learning goals and a district mission, and stage two is focused on developing common understanding of the mission and goals, and on gaining commitment to action in support of student learning in all goal areas. Olson indicates that it is important to have people with differing perspectives and with differing characteristics, such as race, ethnicity, gender, and income level, in the process (the "vertical slice" concept is one important aspect of *Creating the Future*, a district-level strategic improvement process developed by the Northwest Regional Educational Laboratory).

Bringing large groups of people together in a "summit" on student learning goals is another approach, described here by **Gary E. Dunkleberger**. Through this three-phase process, the Carroll County Public Schools in Maryland involved many stakeholders (including professionals, parents, and business representatives) and reached consensus on outcomes and indicators in one month's time. According to the author, the outcomes from the summit process represented the community view of what students should learn better than any set of outcomes developed by a committee.

A third process for defining outcomes is described by **Terry Mazany and Pamala Noli**. In a description and reflection format, Mazany and Noli describe an iterative process in which developers go back and forth between day-to-day student learning experiences and assessment activities. Each iteration builds increased understanding of the intended outcome, learning experience, and assessment activities. When the process involves a community of learners, the resulting outcomes become shared by all who participate.

## Personal Meaning Leads to Commitment

Each of the processes described above leads people in knowledge creation experiences that result in sets of student learning goals that have deep meaning for all the people to which the schooling process is important—students, parents, teachers, other staff, and community members at large. All key stakeholders must see clearly the end result of the years of schooling and be willing to work toward learning success for all students.

Developing shared meaning is different from the more traditional approach to goal definition in which "experts" develop the meanings in think-tank situations, describe the meanings in writing, and communicate the meanings to others. People develop shared meaning through a learning process that involves all key stakeholders in creating and recreating the meaning of the student learning goals. Shared meaning grows over time and through many interactions among stakeholders. Shared understanding rests in the minds and hearts of the people involved in the process. Through the process of developing shared understanding, people gain a personal understanding of what is important and actively commit themselves to promoting learning for all students. Products such as written descriptions are of secondary importance. The products are the symbols of progress that are needed to keep restructuring process invigorated, but they are far from sufficient.

In an excerpt from his book *The Fifth Discipline: The Art and Practice of the Learning Organization*, **Peter Senge** describes the discipline of building shared vision. According to Senge, the members of an organization such as a school or school district will pull together only when there is an overarching goal that people truly want to achieve. In the case of schools, broad statements of student learning goals can provide the vision to which people can commit themselves. Senge describes how shared vision develops from personal vision. When people have a personal vision of what they want to achieve, they commit their time and energy to achieving their vision. Shared vision—in this case agreement on what is important for students to learn to be successful in the 21st century—develops through processes of sharing personal vision, finding points of agreement, and committing to work toward achieving the common aspects of their visions. Over time and through constant dialogue among many individuals, a broad, shared vision emerges. On the other hand, if the vision is developed by a few "leaders" and given to others, the best that will be achieved is compliance with requirements. In this case, when the "leaders" leave, the vision goes and compliance ends.

## Connecting to the Curriculum

In a landmark presentation at the conference marking the 50th anniversary of the Association of Supervision and Curriculum Development, the late **Ernest Boyer** described a far-reaching vision of what students need to be successful in the 21st century. Boyer suggests that the curriculum be organized around "core commonalities," the universal human experiences that are shared by all people and all cultures on the planet. According to Boyer, the points of commonality are the life cycle, birth, growth and death; use of symbols, language, to express feelings and ideas; response to the aesthetic, the arts; placement of ourselves in time and space, geography, astronomy and history; membership in groups and institutions, family, governments, informal structures, and cultures; production, consumption, and conservation; the ecology of the planet; and the search for larger purpose, bringing special meaning to our lives. In his rich descriptions, Boyer implies both what students should learn and how to organize experiences to promote common learning.

### REFERENCES

Senge, P. (1990). *The Fifth Discipline: The Art and Practice of the Learning Organization*. New York: Doubleday/Currency.

Wheatley, M. J. (1992). *Leadership and the New Science: Learning about Organization from an Orderly Universe*. San Francisco, Calif.: Berrett-Koehler Publishers.

# III-2

# Workplace Basics: The Skills Employers Want

ANTHONY P. CARNEVALE
COMMITTEE FOR ECONOMIC DEVELOPMENT, WASHINGTON, D.C.

LEILA J. GAINER
LHW&H, MYRTLE BEACH, SOUTH CAROLINA

ANN S. MELTZER
KRA CORPORATION, SILVER SPRING, MARYLAND

Reprinted with permission from the authors. Copyright © 1990 by the American Society for Training and Development and Jossey-Bass Inc., Publishers. All rights reserved.

# Workplace Basics:

## THE SKILLS EMPLOYERS WANT

*by*
***Anthony P. Carnevale***
***Leila J. Gainer***
***Ann S. Meltzer***

**The American Society for Training & Development**
***William N. Yeomans, President***
***Curtis Plott, Exec. Vice President***

**U.S. Department of Labor**
**Employment & Training Administration**
***Ann McLaughlin, Secretary***
***Roberts T. Jones, Asst. Secretary***

# PART II

## The Skills Employers Want

Many employers say that the most important skills for any employee are the academic triumvirate—reading, writing and computation. With increasing regularity, employers are telling the media, "Give me people who can read, write, and do simple math and I'll train them for the jobs I have available." But probing further, one finds that employers want good basic academic skills *and much more.*

Employer complaints focus on serious deficiencies in areas that include problem solving, personal management, and interpersonal skills. The abilities to conceptualize, organize, and verbalize thoughts, resolve conflicts, and work in teams are increasingly cited as critical.

Their concern is driven by the most compelling of circumstance—economic need. Competitive challenges are forcing employers to adopt an array of competitive strategies that can only be successfully implemented by an innovative and flexible workforce.

Beneath the surface comments about basic skills deficiencies lie employer concerns that they will not be able to achieve their competitive goals with their existing workforce. That they will not be able to successfully integrate new technology or sophisticated production processes. That basic workplace skills deficiencies are beginning to affect their bottom line.

So what are the skills—these basic workplace skills—that employers want? They certainly include basic skills associated with formal schooling. But academic skills such as reading, writing and arithmetic comprise just the tip of the iceberg.

Employers want employees who can learn the particular skills of an available job— who have "learned how to learn."

Employers want employees who will hear the key points that make up a customer's concerns (listening) and who can convey an adequate response (oral communications).

Employers want employees who can think on their feet (problem-solving) and who can come up with innovative solutions when needed (creative thinking).

Employers want employees who have pride in themselves and their potential to be successful (self-esteem); who know how to get things done (goal setting/motivation); and who have some sense of the skills needed to perform well in the workplace (personal and career development).

Employers want employees who can get along with customers, suppliers or co-workers (interpersonal and negotiation skills); who can work with others to achieve a goal (teamwork); who have some sense of where the organization is headed and what they must do to make a contribution (organizational effectiveness); and who can assume responsibility and motivate co-workers when necessary (leadership).

This is a prescription for a well-rounded worker who has acquired a number of discrete skills and who has the capability to acquire more sophisticated skills when necessary.

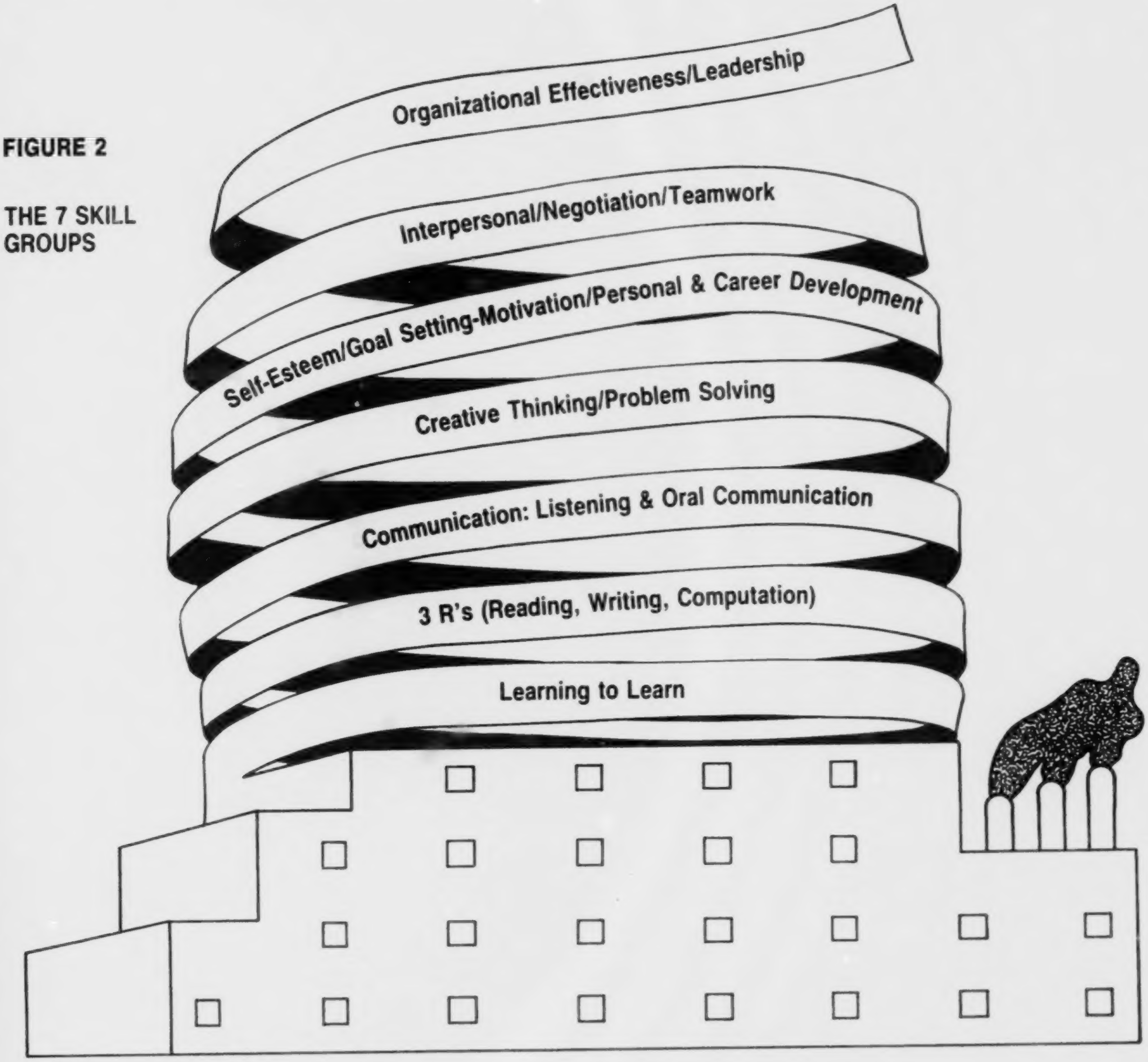
## The Foundation: Knowing How to Learn

The accumulation of knowledge in our society is respected and encouraged. But rarely do we, as children, learn the principles that prepare us to absorb and apply information effectively.

**Knowing how to learn** is the most basic of all skills because it is the key that unlocks future success. Equipped with this skill, an individual

**FIGURE 2**

**THE 7 SKILL GROUPS**


Organizational Effectiveness/Leadership
Interpersonal/Negotiation/Teamwork
Self-Esteem/Goal Setting-Motivation/Personal & Career Development
Creative Thinking/Problem Solving
Communication: Listening & Oral Communication
3 R's (Reading, Writing, Computation)
Learning to Learn

can achieve competency in all other basic workplace skills from reading through leadership. Without this skill, learning is not as rapid nor as efficient and comprehensive.

The first step in providing this skill is to discover how an employee best absorbs information. Identifying a trainee's sensory preference—visual, auditory, tactile—and designing training that integrates that preference is essential. Training in this skill also means exposing the employee to various learning strategies and analytical approaches and providing instruction on how best to apply these tools.

Often, however, just being aware of one's own preferred learning approaches and becoming exposed to various learning strategies will inspire dramatic improvements in learning.

Learning is now a fact of life in the workplace. Even routine jobs are evolving as the demands of the workplace expand. Competitive pressures compel employers to shift employees between jobs and responsibilities, putting a premium on the ability to absorb, process, and apply new information quickly and effectively. The complexity, amount, and availability of information compounds the issue.

From the employer's perspective, an employee who knows how to learn is more cost-effective because time and resources spent on training can be reduced. Employers recognize that long-term relationships with employees are the most cost-effective; therefore, employee ability to adapt to company needs through retraining programs becomes crucial as technology creates shifts in job market demand and job content. Employers see the skill of knowing how to learn as the key to retraining efforts.

But most important, the employees who know how to learn can greatly assist an employer in meeting its strategic goals and competitive challenges, by more efficiently applying new knowledge to job duties and tasks.

Productivity, innovation, and competitiveness all depend on developing the learning capability of the workforce. Machinery and processes are transferable from company to company and country to country. But it is the application of human resources to technology and systems that provides the competitive edge.

## Competence: Reading, Writing and Computation

Basic academic skills—reading, writing and computation- have long been revered as the keys to success in society and the workplace. In theory these skills have been essential, but in practice workers have often succeeded because of "a strong back and willing hands."

The workplace of the past was one where those with limited academic achievements could succeed. Jobs often required going through the motions of a regularized process or repetitive interaction with machines. In that workplace, illiteracy and innumeracy could be hidden or ignored.

But today's workplace is one that increasingly involves interaction with sophisticated, computerized machinery that requires good reading skills for efficient use. The introduction of approaches such as statistical process control (SPC) demand higher mathematical skills. And writing is frequently the first step in communicating with customers, interacting with machines, documenting competitive transactions, or successfully moving new ideas into the workplace.

America is fortunate in that the vast majority of our workers are literate and numerate. But frequently, employees cannot use these skills effectively in the workplace. Sometimes it is because they are "rusty" and are called upon to use mathematical principles they haven't used for 20 years or so. Sometimes it is because skills must be used in a different context than originally learned. Sometimes, the base knowledge is there but there is no understanding of how to expand and apply it. All of these problems create the need for employer-provided training in reading, writing, and computation.

The challenge of providing traditional academic skills in the workplace is one that can be met by using an applied approach to basic workplace skills training, using or adapting materials and concepts that are job-based.

This requires an awareness that sometimes the intended use of a skill makes that skill more complex. For example, traditional classroom reading instruction is designed to teach discrete **reading** skills, in isolation, for the purpose of increasing a student's ability to follow directions or internalize data for future recall. For this reason, even some students who master reading in an academic setting are not prepared for the way they will use the skill in the workplace.

Reading tasks on the jobs are different. They require the reader to be analytical, to summarize information, and to monitor one's own comprehension of the reading task. This is an interpretive approach that requires the reader to have active involvement with the reading task. From the employer perspective, basic training in reading should focus on teaching reading processes for locating information and for using higher level thinking strategies to solve problems.

Similarly, the traditional classroom approach to teaching **writing** focuses on description (creative writing) or on the articulation of learned responses concerning facts and events (essays). Workplace writing relies on analysis, conceptualization, synthesis and distillation of information, and clear, succinct articulation of points and proposals.

In schools, **mathematical** concepts are presented sequentially, beginning with fundamental concepts like addition and subtraction and working through to higher level operations such as geometry and trigonometry. Traditional math education is a group activity with the instructor and the group focusing on sample problems presented by the instructor. This is followed by independent practice, repeated drills, and finally standardized tests to measure proficiency *level*.

Workplace math skills are taught contextually to reflect their actual use on the job; instructional materials simulate specific job tasks. Building on the prior math knowledge of the learner and emphasizing problem identification, reasoning, estimation, and problem-solving, this approach has been shown to produce the quickest, most effective results in employee performance.

Most employers today cannot compete successfully without a workforce that has sound basic academic skills. Workers spend an average of one and one-half to two hours per workday engaged in reading forms, charts, graphs, manuals, computer terminals, and so on. Writing remains the primary form of communication for transmitting policies, procedures, and concepts. Computation is used daily to conduct inventories, report on production levels, measure machine parts or specifications, and so on.

Deficiencies in such basic workplace skills create barriers that impair an employer's ability to meet strategic goals and to be competitive. They are reflected in productivity decline, increased accident rates, costly production errors, and the inability to effect critical job retraining.

# Communication: Listening and Oral Communication

Reading and writing are essential communication tools, but it is through listening and speaking that we interact most frequently. The average person spends 8.4 percent of communications time writing, 13.3 percent reading, 23 percent speaking, and 55 percent listening.

Workers spend most of their day in some form of communication. They communicate with each other about procedures and problems and they relay information to and receive it from customers. Success on the job is linked to good communications skills. In fact, recent studies have indicated that only job knowledge ranks above communications skills as a factor for workplace success.

Ironically, American schools offer scant instruction in oral communication or listening. While extensive formal training in reading and writing is provided, instruction in speaking skills is generally provided as an elective course linked to drama, debating, and so on. Virtually no training in listening is available.

Instruction in effective **oral communication** skills rests on providing workers with an understanding of the importance of voice inflection and body language when communicating. The styles that people use when communicating affect how they are perceived and what is heard. Workplace training in oral communication simulates actual workplace circumstances and instructs employees in recognizing their dominant styles of communication and how they are manifested, understanding and valuing communication approaches that are different in style from their own, and adjusting their dominant style to someone whose style differs.

**Listening** skills curricula should also simulate actual workplace experiences and assist trainees to understand how listening style affects the transmission and receipt of information. Instruction should focus on the five listening skills that are critical for workplace success: listening for content; listening to conversations; listening for long-term contexts; listening for emotional meaning; and listening to follow directions.

Employees who lack proficiency in oral communication and listening skills are handicapped as to their learning and communicating abilities, and their personal and professional development. Business leaders estimate that deficiencies in these skills cost employers millions each year in lost productivity and errors.

Communication is central to the smooth operation of a competitive venture. Communications skills are at the heart of getting and keeping customers. Pitching innovation, contributing to quality circles, resolving conflict and providing meaningful feedback all hinge on oral communication and listening skills.

# Adaptability: Creative Thinking and Problem-Solving

An organization's ability to achieve its strategic objectives often depends on how quickly and effectively it can transcend barriers to improved productivity and competitiveness. These pressures put creative thinking and problem solving at a premium—at all levels of an organization.

**Problem solving** skills include the ability to recognize and define problems, invent and implement solutions, and track and evaluate results. Cognitive skills, group interaction skills, and problem-processing skills are all crucial to successful problem solving. Workplace training programs in problem solving simulate real problems and are keyed to the organization's goals.

New approaches to problem solving, organizational design, or product development all spring from the individual capacity for creative thinking. In the workplace, creative thinking is generally manifested as creative problem solving or creative innovation. Often a group activity, creative problem solving is characterized by effective teamwork, the examination of problems in new ways, and the invention of new solutions to existing problems. Either an individual or group activity, creative innovation refers to the development of new activities that expand markets, and improve such elements as productivity.

Training in **creative thinking** is designed to expand the thinking processes of trainees, to allow trainees to escape from logical and sequential thought patterns. For example, exercises to promote divergent thinking (finding connections between seemingly unrelated ideas) might be conducted. Most training in creative thinking involves problem solving, personality awareness and development, and group teambuilding activities.

An organization's ability to achieve its strategic objectives often depends on the problem-solving and creative-thinking skills of its workforce. Unresolved problems create dysfunctional relationships in the workplace. Ultimately, they become impediments to flexibility and to dealing with strategic change in an open-ended and creative way. Creative solutions help the organization to move forward toward its strategic goals.

# Personal Management: Self Esteem, Goal Setting/Motivation, and Personal/Career Development

An individual's effectiveness in the workplace can be linked directly to positive self-esteem and successful personal management. A good self-image means the employee takes pride in his or her work. The ability to set goals and meet them will become evident as production quotas are exceeded or deadlines met. Solid personal and career development skills are apparent in efficient integration of new technology or processes, creative thinking, high productivity, and a pursuit of skills enhancement through training or education.

In the past, employers viewed employees with solid occupation-specific skills as having all "the ight stuff" for success on the job. But today workers are increasingly called upon to make decisions at the point of production or point of sale and to display good interpersonal skills when they work in teams or with customers. The confidence that engenders success in these areas springs from a positive sense of self worth.

Ironically, just as employers are recognizing the need for employees with positive self esteem, they are faced with a shrinking entry-level labor pool which will propel them toward selecting workers from populations that have been buffeted by poverty, circumstance, and environment— workers who frequently have low self-esteem. And at the same time, the frustrations of business today—frequent mergers, downsizing, lay-offs—can erode the self-esteem of established workers who feel powerless or unable to progress toward their career goals.

Workplace training to enhance **self-esteem** is at the core of training to expand job-specific occupational skills or provide remediation. Key elements of self-esteem training include assisting employees to recognize their current skills; be aware of their impact on others; understand their emotional set points and abilities to cope with stress, change, criticism, and so on; and deal with their own limits by recognizing the need for and seeking new information to apply to problems and construct solutions.

For an employer to succeed in the marketplace, employees must be motivated; they must possess the ability to set and to meet reasonable goals. From the shop floor to the salesroom, the motivated employee comes out on top.

Poor performance can often be linked to deficiencies in these skills (although sometimes the root causes are mitigating circumstances such as low pay or poor working conditions). Individual employees' lack of motivation or goal-setting skills can produce an organizational undercurrent of repeated errors, absenteeism, and quality problems, or it can construct barriers along the path to change.

Training in **motivation/goal setting** is rooted in techniques for setting and meeting defined objectives, including recognizing signposts of incremental success along the way toward meeting a goal. Like self-esteem training, it focuses on self-awareness and adaptability as keys to dealing with things or people. It also emphasizes self-direction and organizational savvy.

**Personal and Career Development** skills rest on a foundation of good self-esteem and motivation. Employees who exhibit these skills increase their value in the workplace and in the employment marketplace. Employers value these skills in employees because they usually indicate successful job transitions and effective training experiences.

Unfortunately, the educational system provides little formal training in defining career direction and identifing the education and training needed to achieve career goals. Consequently, many people enter the workforce with little understanding of these skills and react to job opportunities as they surface. This patchwork-quilt approach to career development, which worked well in a time when employees could reasonably expect a long-term career with one employer, is no longer viable.

Training in personal and career development skills includes providing employees with techniques for understanding and expanding their skills inventories, career planning, and career management. Goal setting is important as is structuring individual career progression models that explore the training and educational preparation needed to meet career goals. Importantly, training should be complemented by on-going counseling and by the organization's clearly articulated support of potential career paths for individual jobs.

Organizations have traditionally viewed the skills of self-esteem, goal setting/motivation, and personal/career development as those an individual should acquire outside the workplace. But the demands of today's evolving workplace are influencing employers to recognize that they must look beyond providing occupation-specific skills. They must include training in the more intangible skills that together make a productive and competitive workforce.

# Group Effectiveness: Interpersonal Skills, Negotiation, and Teamwork

In the past two decades, there has been a tremendous increase in the use of teams in the workplace. The team approach has been linked conclusively to higher productivity and product quality, as well as to increased quality of worklife. Change strategies are usually dependent upon the ability of employees to pull together and refocus on the new common goal.

Whenever people work together, successful interaction depends upon effective interpersonal skills, focused negotiation, and a sense of group purpose. The quality of these three factors defines and controls working relationships.

**Interpersonal skills** training is directed toward assisting the employee to recognize and improve the ability to judge and balance appropriate behavior, cope with undesirable behavior in others, absorb stress, deal with ambiguity, listen, inspire confidence in others, structure social interaction, share responsibility, and interact easily with others.

Such skills are essential to successful **negotiation**. Conflicts, both major and minor, are a fact of worklife. They can sap productivity and short-circuit strategic plans.

The key to diffusing potential conflict situations is to enhance employee negotiating skills at all levels.

Training in this skill includes techniques for separating people from the problem, focusing on interests not positions, inventing options for mutual gain, and insisting on the use of objective criteria. It also relies on a sound base of interpersonal skills and a clear understanding of the approach demanded by the circumstance.

Interpersonal and negotiation skills are the cornerstones of successful **teamwork**. Teams are organized in the workplace so that appropriate talents and skills can be directed through group effort to accomplish vital tasks and goals. This pooling of resources, however, frequently requires team members to have an array of skills that individual or routine jobs do not demand.

Quality teamwork results when team members know how to recognize and cope with various and unique personalities and when each has a sense of the cultures and approaches that other team members represent. Team members also need an understanding of group dynamics, which evolve and change as the team approaches its goal. Lastly, team members must be aware of the technical skills that fellow members have and how those skills can be applied.

Teamwork can only occur when team members provide and receive feedback in a focused manner. Individuals gather and process information in personalized ways; good teamwork calls for the recognition and use of certain valuable differences between members of the team.

Training in teamwork is crafted to capture the essential elements of building team relationships. The major objective is to develop an inventory of skills and attitudes that can be applied successfully in the workplace to resolve problems and foster innovation.

The strategic relevance of interpersonal, negotiation, and teamwork skills is evident. They are basic tools for achieving the flexibility and adaptability that America's workforce must have in order to remain competitive.

# Influence: Organizational Effectiveness and Leadership

To be effective in an organization, employees need a sense of the workings of the organization and how their actions affect organizational and strategic objectives. Skilled in scoping out the forces and factors that interfere with the employer's ability to accomplish its tasks, the worker can be a master problem-solver, an innovator, and a team-builder.

Organizations are a tapestry of explicit and implicit power structures. In the explicit structure, leadership is conferred and represented by title and authority. In the implicit structure, leadership is a delicately sculpted image built by cultivating the respect of peers, and projecting a sense of reliability, goal orientation, and vision.

Both organizational effectiveness and leadership skills are basics for success in tomorrow's workplace. Those who have these skills can help employers create the conditions for achieving goals and for succeeding in the marketplace.

Basic training in **organizational effectiveness** is geared toward providing the trainee with an understanding of what organizations are, why they exist, and how one can navigate the complex social waters of varying types of organizations. Once armed with this basic framework, the trainee is exposed to the organizational culture of his or her employer, its goals, values, culture and traditional modes of operation. Lastly, training in skills that make the employee a fully functioning member of the organization - interpersonal, communications, and group dynamics skills - completes the picture.

Organizational effectiveness skills are the building blocks for leadership. Unaccompanied by them, leadership skills can be misplaced and even counterproductive.

At its most elementary level, **leadership** means that a person can influence others to act in a certain way. Every person may need at times

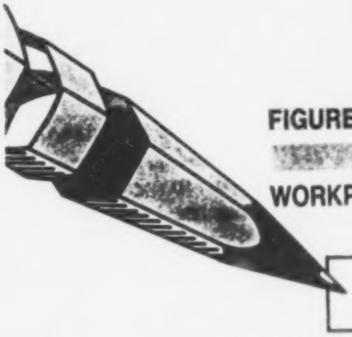


**FIGURE 3**

**WORKPLACE BASICS: A CHECKLIST**

☐ Learning to Learn

☐ 3 R's (Reading, Writing, Computation)

☐ Communication: Listening & Oral Communication

☐ Creative Thinking/Problem Solving

☐ Self-Esteem/Goal Setting-Motivation/Personal & Career Development

☐ Interpersonal/Negotiation/Teamwork

☐ Organizational Effectiveness/Leadership

to influence his or her work group and to provide a vision of what the organization as a whole or the specific task at hand requires. Leadership skills are necessary, therefore, for every level of the organization from the chief executive to the line worker.

Leadership skills training for the average employee includes the following components: how to understand the organization's strategies and tactics for achieving goals; leadership as an exchange process between leaders and followers; situational approaches for the task-centered leader; strategies for sound decision-making; developing and communicating a vision; influencing the behavior of others; and the importance of projecting emotional stability.

To compete in world markets, employers need employees who can operate effectively within the parameters of their organization, assume responsibility willingly, and motivate coworkers toward exemplary performance. These skills—once identified with the "fast track"—are now basic ingredients.

# III-3

## Civic Education for Responsible Citizens

ERNEST L. BOYER
THE CARNEGIE FOUNDATION FOR THE ADVANCEMENT OF TEACHING

First published in *Educational Leadership* 48, 3 (November 1990): 4–7. Reprinted with permission from the author.

ERNEST L. BOYER

# Civic Education for Responsible Citizens

To further our tradition of government by the people and to preserve our nation's well-being, education for citizenship is not an option—it's what schools are for.

IN CONGRESS, July 4
The unanimous
America

Thirty years ago, the Soviets' launch of Sputnik touched off this country's last big push for school reform. We wanted scientists to compete in the space race and, given the Cold War climate of the day, children were bundled off into physics and chemistry classes, slide rules in hand, to assure that America would remain militarily secure.

This time, the reform movement has been driven almost exclusively by economic threats and by the growing recognition that without good schools the financial well-being of the nation is imperiled. Education and the economy are inextricably interlocked, and it is now widely agreed that if the United States is to regain its competitive advantage, mind power is the key.

No one denies that America's work force must get better. But there's another imperative. While economic purposes are being vigorously pursued, civic priorities also must be affirmed. Indeed, unless we find better ways to educate ourselves as *citizens*, America runs the risk of drifting unwittingly into a new kind of dark age, a time when specialists control the decision-making process and citizens will be forced to make critical decisions, not on the basis of what they know, but on the basis of blind belief in so-called "experts."

### The Decline in Civic Education

From the earliest days of the Republic, schools accepted the obligation to participate in the building of a nation. Benjamin Rush, writing in 1786, insisted that youth should be educated to "watch for the state as if its liberties depended on [their] vigilance alone." And when Jefferson was asked if mass opinion could be trusted, he responded:

> I know no safe depository of the ultimate powers of the society but the people themselves, and if we think them not enlightened enough to exercise their control with a wholesome discretion, the remedy is not to take it from them, but to *inform their discretion* by education (italics added).[1]

The establishment of a network of common schools in the 19th century was, in large measure, an attempt to strengthen democracy. The push for universal education was driven less by individual gain than by the desire to promote the social and civic advancement of the nation, based on the belief that we had, in this country, a rich heritage to be shared and a vision of participatory government to be sustained and strengthened.

**Civic education means helping students develop responsible ways of thinking, believing, and *acting*.**

The problem is that today the Jeffersonian vision of grass roots democracy fueled by education increasingly is viewed as utopian, and what's especially disturbing is that the school reform movement of the 1980s has paid insufficient attention to educating students about our nation's history and institutions.

An indication of the problem's proportion is found in a recent National Assessment of Educational Progress survey, which found that only 53 percent of 8th graders and just 57 percent of high school seniors were aware that the Declaration of Independence affirmed the right to life, liberty, and the pursuit of happiness. And only 44 percent of the 8th graders and slightly more than half the 12th graders knew that the right to religious freedom is found in the amendments to the Constitution.[2]

Adding to the worry is a critique of civics textbooks issued by James Carroll of the Brookings Institution, who concluded:

> many of the books are largely disembodied expositions of principles and facts, lacking the passion of the conflicts that infuse politics and government with meaning and significance. The student is asked only to master knowledge of the subject rather than to put his knowledge to use. Thus, the participatory side, the side that requires the individual to analyze democratic values, processes, and choices, is largely ignored.[3]

Little wonder that voter participation is falling off. Turnout in the 1988 presidential election campaign was only 50 percent, the lowest in 64 years. A recent report by the Markle Commission on the Media and the Electorate noted that "American voters today do not seem to understand their rightful place in the operation of American democracy" and "act as if presidential elections belong to some[illegible] else."[4]

The good news is th[illegible]orts to strengthen citizenship e[illegible] are emerging. For several ye[illegible] the National Commission on Social Studies has brought together school and college leaders to define priorities for the study of government and other social institutions. Further, the Council for the Advancement of Citizenship and the Center for Civic Education have been working collaboratively on a curriculum framework called "CIVITAS," which stresses critical thinking and participation skills. In addition, efforts in several states, California among them, are seeking to strengthen the social science curriculum.

But, with all of this, there remains a widespread concern that education for citizenship has been neglected, and for those who care about government "by the people," the decline in civic education cannot go unchallenged. In a world where human survival is at stake, ignorance about public issues is not an acceptable alternative. What we urgently need today are groups of well-informed, caring individuals who band together in the spirit of community to learn from one another and participate, as citizens, in the democratic process. But how should we proceed?

### What Civic Education Must Provide

To help students become responsible citizens, civic education programs should have several important characteristics.

*Civic education is concerned, first, with communication.* The work of democracy is carried on through thoughtful discourse—town meetings, city councils, study groups, informal conversation, the television screen—and citizenship training, if it means anything at all, it means teaching students to think critically, listen with discernment, and communicate with power and precision.

If students learn to listen and speak more carefully, as well as read and write, they will not only be civically empowered; they also will know how to distinguish between the authentic and the fraudulent in human discourse. Indeed, the destiny of this country may be threatened not so much by weapons systems but by the inclination of public officials to send messages that obscure truth.

*Civic education also must provide students with a core of basic knowledge regarding social issues and institutions to give their understanding of democracy perspective.* In the Carnegie report *High School*, we recommended that all students study American history and government. We urged that they be introduced to political thinkers, from Plato and Locke to John Adams and James Madison, and that they study the Declaration of Independence, the Constitution, and the *Federalist Papers*. Equally important, students should learn about the heritage of other cultures as they prepare to take their places in a world of diverse peoples whose destinies have become intertwined.

*Further, civics classrooms should be active, not passive, places.* Unfortunately, teaching about government and its functions all too often has focused on textbooks, without involving students sufficiently in the *processes* of decision making. The National Assessment of Educational Progress found that textbook reading assignments were, by far, the most common method of instruction in civics classrooms. Other activities, such as writing papers and working on group projects, were less prevalent.[5]

When 12th graders were asked how often they participated in mock trials, imitation elections, or governmental bodies, more than half said they had never done so. And yet the National Assessment reported that students exposed to such learning performed better than peers who had only occasionally, or never, participated in these activities.[6]

I'm suggesting that for civic education to come to life, theories of government must be tested. Students should become knowledgeable about contemporary issues and be asked to thoughtfully weigh the options. For example, each student might take a contested issue currently being considered by the Congress, a state legislature, or local government body and report in depth on the history of the problem, define points of tension, and propose plausible solutions. The classroom itself should become a town meeting.

*We must deal thoughtfully with our deepest differences.* All of this sounds fine in principle, but what about issues that appear too hot to handle? The harsh truth is that the national consensus on many issues seems to have eroded. Society today is characterized by such divisiveness that consensus seems almost impossible, and this puts the nation's schools squarely in the middle. Educators are often confused—even abused—if they try to examine touchy social problems and to help students debate what constitutes the common good. Yet to ignore controversial issues is to offer students an incomplete education, an incapacity to think carefully about life's most important concerns. I remain convinced that even in matters where society is sharply divided, schools have an especially important role to play, one that goes beyond silence or the extension of the status quo.

If we hope to make progress toward resolving deep conflicts in the culture, we must encourage open and sensitive classroom discussion about choices, even in such controversial areas as sex, drugs, cultural differences, and religious beliefs. Finding a way to deal thoughtfully with our deepest differences is perhaps the greatest challenge citizenship education now confronts. And in the guidance of such inquiry, teachers must be trusted.

Civic education must be taught in other ways as well. *Students, while they are in school, are members of an institution, and they should understand how it works and participate, within the school itself, in decisions that affect their lives—just as they will be asked to do later on.* In *Classroom Life as Civic Education*, David Bricker makes the essential point that much of what young people learn about citizenship comes to them "indirectly as they draw out ideas about how people should conduct themselves in public, from the ways their teachers manage classroom life."[7]

If we postpone such involvement for students while they're young, there is a good probability that it will be deferred for a lifetime. Clearly, citizenship is not something to be deferred. It should be demonstrated in every institution in which the student is involved, especially at school.

*Finally, education for citizenship means helping students make connections between what they learn and how they live.* During our study of the American high school, we found that too many young people feel unwanted, unneeded, and unconnected to the larger world. Further, there is a serious gap between the young and the old in our society, an intergenerational separation in which youth and their elders are not seriously engaged in common discourse.

I'm convinced that students, as an essential part of civic education, need to understand that learning is for living, that education means developing the capacity to make judgments, form convictions, and act boldly on values held. In response to this challenge, we proposed in *High School* a new Carnegie unit, one based on service. The idea is that every student should be asked to volunteer his or her time at retirement villages, daycare centers, or youth camps, or to tutor other kids at school.

But there's a caution here. While such projects can generate within students a sense of worth, they must be viewed as part of the educational experience and not just an after-school activity. Specifically, service projects should include a written evaluation by the student, linking community activity to classroom theory.

I'm suggesting that civic education, by its very nature, means helping students confront social and ethical concerns and apply what they have learned. We must help them understand that not all choices—in thought and action—are equally valid. Such an education does not dictate solutions or suggest that there are simple answers for every complicated question. Rather, it means helping students develop responsible ways of thinking, believing, and *acting*.

### Today's Window of Opportunity

We stand at a strategic time in the history of public education. Interest in schools has run high during the past six years. But public attention cannot be indefinitely sustained. Today's window of opportunity hinges to a great extent on the ability of the schools to pursue, not just the economic, but the civic ends of education.

When all is said and done, the nation's schools should encourage each student to develop the capacity to judge wisely in matters of life and conduct. Time must be taken in classrooms to explore ambiguities and reflect on the consequential social issues of our time. The goal is not to indoctrinate students but to provide a climate in which civic choices can be thoughtfully examined and convictions formed. These are the characteristics by which, ultimately, the quality of public education must be measured. ☐

---

[1] T. Jefferson, *Letters*, arranged by W. Whitman, (Eau Claire, Wis.: E.M. Hale, 1940), pp. 338, 339.

[2] The National Assessment of Educational Progress, (1990), *The Civics Report Card*, (Princeton, N.J.: Educational Testing Service), p. 65.

[3] J. D. Carroll et. al., (1987), *We the People: A Review of U.S. Government and Civics Textbooks*, (Washington, D.C.: People for the American Way), p. iv.

[4] R. Rothman, (May 16, 1990), "Panel Urges Education to Revitalize Citizens' Participation," *Education Week*, p. 8.

[5] The National Assessment of Educational Progress, *The Civics Report Card*, p. 87.

[6] Ibid., p. 85.

[7] D. C. Bricker, (1989), *Classroom Life as Civic Education*, (New York: Teachers College Press), p. 1.

# III-4

# What Really Counts in Schools

Elliot W. Eisner
Stanford University

First published in *Educational Leadership* 49, 5 (February 1991): 10–17. Reprinted with permission from the author.

ELLIOT W. EISNER

# What Really Counts in Schools

Our understanding of the mission of schooling must go beyond the merely measurable to a consideration of more profound purposes.

Deciding what really counts in our schools is the first order of our educational mission. The second is to create the conditions that are consistent with that mission. There is no shortage of opinion about our mission. We hear from the highest offices in our land that American schools are in a dismal state. We hear from others that they are in a state of crisis. We read in the mass media about our lowly position, not in the arms race, but in the "education race." Education has become the front line in our quest for international supremacy. For many, there is no mistaking it: Education is not only a business, it is a competitive race.

We are told that we must get our house in order, in order to continue to reside in our house—others are waiting in the wings. The mark of our success will be our position among other nations on tests that purportedly measure what counts.

I have more than severe reservations about such prescriptions for educational supremacy. Neither supremacy nor the metaphor of the race are, I believe, appropriate for thinking about the kinds of lives we would like our children to lead. A president who, when campaigning for office, spoke of the need for "a kinder and gentler

America" cannot logically view our nation's schools as instruments for beating others. There is in such observations, at the very least, a tinge of inconsistency.

It is time to think about what really counts in schools. I think that our analysis ought to penetrate the simplistic, the merely measurable, and the superficial. This article is my effort to make that penetration.

## A Starting Point

I start with certain premises about human nature and the mission of education. I start with the premise that when children come into the world, they do not possess minds. People acquire minds during the course of their lifetimes; the task is not completed when students finish the 12th grade. Children *do* come into the world with brains; brains are biological. Minds are cultural. Minds are forms of cultural achievement, and the kinds of minds that children come to own depend, in large measure, on the kinds of experiences they are afforded during the course of their lifetimes.

*One important aim of schooling should be to create a climate that evokes children's sense of wonder and inspires their imaginations to soar.*

We all know what a *culture* is. In the biological sense, a culture is a place for growing things. Schools are cultures for growing minds. The major instruments that we use to create minds are the kinds of programs we offer and the quality of teaching we provide. Like the systole and diastole of the beating heart, curriculum and teaching are the most fundamental aspects of our schools. Decisions that we make about what to include and what to exclude from the programs we offer are of the foremost importance. But programs, no matter how well-conceived, must always be mediated if they are to influence the lives of those with whom we work. This process of mediation, at its best, is an artistic activity. We call it *teaching*. When teachers transform the inevitably limited and schematic conceptions of school programs into the kinds of activities that genuinely engage students, when they create the environments that open up new vistas and provide for deep satisfactions, they make a difference in the lives that children lead. No curriculum teaches itself, it always must be mediated, and teaching is the fundamental mediator.

Teaching, however, as a form of human mediation, is not the only important influence on children. How we organize the "envelope" within which teaching and curriculum activities occur also matters. That is, how schools are structured, the kind of values that pervade them, the ways in which roles are defined and assessments made are a part of the living context in which both teachers and students must function. Make no mistake about it, these pervasive organizational features define the kind of place that schools are; they also teach. Dreeben (1968) refers to these features as the non-pedagogical aspects of instruction.

## Six Aims That Count in Schools

What really counts in school, from my perspective, may not appear on anyone's list of national educational imperatives. Yet, I believe they are imperative. Let me share them with you.

### *The Journey Is the Reward*

First, what really counts in schools is teaching children that the exploration of ideas is sometimes difficult, often exciting, and occasionally fun.

I know that such a prosaic aspiration for educational practice as a fundamental aim of education must seem simple. In one sense, it *is* simple. It is certainly an aim that is simple to articulate. It is not so simple to achieve. But it is of the utmost importance educationally.

We sometimes seem to believe that the importance of what we do in school is determined by how well children do on tasks they will encounter later in schools or in colleges or universities. I take issue with such a conception. The major goals of schooling are not realized by performances on tasks defined in classrooms or within schools. The important effects of schools are located in the kinds of lives that children lead outside school and the kinds of satisfactions they pursue there. In research terms, the major dependent variables of schooling are not scores on standardized achievement tests, whether norm- or criterion-referenced: they are the kinds of ideas children are willing to explore on their own, the kinds of critical skills they are able to employ on tasks outside classrooms, and the strength of their curiosity in pursuing the issues they will inevitably encounter in the course of their lives. Indeed, a much better index for school achievement than standardized achievement test scores is the level and quality of the conversations children engage in away from their classrooms.

The reason it is so important for youngsters to enjoy what they study in school is because without such satisfactions, the likelihood that they will pursue their studies outside our institutions is small.

*Continued on p. 14*

*Photographs by Kate Leone*

*Imagination, a distinctly human trait, needs to be valued by the schools, not inhibited*

*Continued from p. 11*

There are, it is said, three reasons that motivate people to do things. The first of these is to secure satisfactions that are intrinsic to the activity itself. Both sex and play are paradigm cases of such satisfactions. In both, activities are pursued for the intrinsic satisfactions that they yield. That is why, I think, intellectual activity at its highest level is often associated with play. When children enjoy playing with ideas and dealing with problems in a playful way, when they secure the kinds of satisfactions that they get from activities that they choose, they are engaged in the kind of activity I am talking about.

A second reason for doing something is not because the activity itself is pleasurable, but because the results are. Most people like to have a clean kitchen, but I suspect few of us enjoy kitchen-cleaning chores. Many people work at jobs whose processes they don't enjoy, but they do take pride and satisfaction in the work once done.

A third reason for doing something is not because the process is satisfying, nor because the results are, but because one likes the payoff in the form of a paycheck. Too many people, lamented Hannah Arendt (1958), are engaged in this kind of activity as a way of securing a livelihood. For Arendt, such activity constitutes labor. Work, by contrast, is related to the first and second forms of motivation.

I happen to think that in our schools many children are inadvertently motivated by extrinsic reward systems that teach them that the important reason for engaging in an activity is to secure a reward that typically has little or nothing to do with the activity itself. We call such motivators "stars," "brownie points," "grades," and "honor classes." What will happen to those students when those motivators are no longer available? Will they take the journey? They will, if in the context of schools and classrooms they discover that the journey is the reward.

## Formulating Questions, Seeking Answers

A second aim that really counts in schools is helping youngsters learn how to formulate their own problems and how to design the tactics and strategies to solve them. We often talk in education about the importance of problem-solving skills, and clearly the ability to solve problems is important. But the most difficult of intellectual tasks is not primarily in solving problems but in formulating the questions that give rise to them. We place great emphasis in our schools on setting tasks for students and expecting them to pursue them. We call these tasks "assignments." They emanate from teachers and are directed to students. The child's job is not to question why, but, as they say, to do or die.

Some writers (Apple 1982) have referred to this process as a process of ***deskilling***. What is deskilled when such a process dominates classrooms and schools is the child's ability to assume responsibility for the aims as well as the means he or she might employ in dealing with them. Since in the course of life the ability to conceptualize what is problematic and to formulate interesting questions is fundamental in maintaining one's intellectual autonomy, it does not seem unreasonable that schools should be places in which students could practice such skills. This would require giving children an opportunity to formulate their own aims, to conceptualize their own problems, and to design the ways in which such problems might be addressed. It means that finding out what students want to know about *after teaching* is as important as determining whether they possess the answers we hoped they would. The kinds of questions students can formulate is as important an educational outcome—perhaps even more so—as how well they can converge upon the correct answers that populate our textbooks.

Providing school programs that make it possible for students to formulate their own problems increasingly as they get older and to work on tasks related to them would, undoubtedly, lead to increased heterogeneity in the classroom. I am fully aware that it is neither feasible nor desirable for such activities to be all that is provided in our curriculums. There does need to be some common fare and some effort at cooperative group activity. Every child going off on his or her own track in an utterly independent way is not what I have in mind. But what I do envision is providing children with opportunities to learn how to formulate their own goals, questions, and problems. Providing such opportunities will, of course, make the process of evaluation more complicated and difficult. However, our educational aspirations should not be defined by the current limits of our testing technology. We need to reach for more.

## Multiple Forms of Literacy

A third aim that counts in schools is the development in the young of multiple forms of literacy. Literacy is typically thought of as the ability to read and write. Literacy surely includes such skills, but I conceptualize it in broader ways (Eisner 1985). Literacy is, broadly speaking, the ability to encode or decode meaning in any of the forms used in the culture to represent meaning. Aristotle observed long ago that "man by nature seeks to know." The ***knowing*** that Aristotle talked about is embedded in human experience, and human experience, in turn, is influenced by the forms of representation that we can meaningfully employ. School programs ought to develop literacy, that is, the ability to secure meaning, from the wide range of forms that are used in culture to express meaning. This surely includes far more than the literal use of language or the ability to write precise "standard" English. Language itself takes many forms. There is the literal use of language, the literary use, the figurative, the poetic, and the vernacular use of language. Language is diverse: the kinds of meaning that can be represented and secured in poetry, for example, are simply unavailable to those limited to prose.

Similarly, the meanings that are engendered through choreography, through music, and through the visual arts are unique or special to their forms. There are some meanings that can be grasped through visual form that cannot be described in language or in quantitative form. Toshiba, IBM, Apple, and other computer companies have long recognized that spreadsheets, pie charts, scattergrams, and visuals in living color increase the meaningfulness of some kinds of information; they know that the way in which we construct meaning depends upon the way in which forms of representation are configured. Not everything that we want to say can be said in language. Not everything that we want to convey can be reported in number. The moral here for school programs is clear: those that neglect or marginalize the fine arts, for example, embrace an educational policy guaranteed to graduate students who are semi-literate. The great ceramic figures of the Zapotec, the bronze vessels of the Han Dynasty, the egg tempera madonnas of 14th century Italy, the Abstract Expressionism of the 1950s will be for our students other people's pleasures if they are denied the opportunity to learn how to see and understand what these works make possible.

At present, American schools are embarked on a set of educational priorities that are both narrow and shortsighted. We think about literacy in the tightest, most constipated of terms. We need a more generous conception of what it means to know and a wider conception of the sources of human understanding. The poet, the painter, the composer, the playwright, as well as the physicist, the chemist, the botanist, the astronomer have something to teach us. Paying adequate attention to such forms of understanding in schools is the best way to make them a meaningful part of our students' intellectual lives.

**The imagination is, fundamentally, an important dimension of human consciousness and, at bottom, the engine of cultural and social progress.**

### *In Praise of Wonder, Imagination*

A fourth aim that really counts in schools is teaching the young the importance of wonder. So much of what we teach in schools seems to undermine the importance of wonder. Indeed, wondering, even daydreaming has been regarded as an "off-task" activity. The tacit image of pedagogical virtue conveyed by so much of our educational literature is an image of youngsters with their heads buried in workbooks who never come up for air. Academic engaged time, by itself, is no virtue. Whether being academically engaged is or is not a virtue depends, at minimum, on the nature of the task in which one is engaged. Often times daydreaming and wonderment are forms of respite from activities that children find meaningless and distasteful. They are means children use to maintain their psychological equilibrium.

But more than the satisfactions afforded by escape from meaninglessness are the potential worlds that wonder makes possible. The imagination is, fundamentally, an important dimension of human consciousness and, at bottom, the engine of cultural and social progress. It is a resource distinctive to our species. Some people have argued that bees are great architects and proceed to support their case by pointing out that bees use a material that is soft and malleable to build not only beautiful architectural structures of great proportion and precision, but structures of great strength as well.

The fact of the matter is, however, that bees are not great architects. Bees create honeycombs because they can do nothing else. They have been creating these same honeycombs for years now, and I suspect that they will continue to do so in the future. What architects require, and what humans possess, is imagination. Humans create buildings that take the forms they do by virtue of wonderment and by the strength of imagination. When bees are able to create Baroque beehives I might change my mind, but until then I am unwilling to regard bees as creative architects.

Wonder and imagination are fundamental not only in architecture, but also in science and in all creative aspects of human affairs. I believe our schools should create the kind of environment and provide the kind of tasks that elicit and develop respect for wonder and stimulate the imagination. One of the features of schooling that dampens the inclination toward wonder and that limits the imagination are the kinds of assessment practices we typically employ. If there's any single lesson that multiple-choice tests teach, it is that for every question there is a single correct answer and for every problem a single correct solution. The correct solution is known by the test-maker or the teacher, and the student soon learns that his or her task is to converge upon the correct one. The tacit message is a message of convergence, of singularity, of homogeneity.

If we created teaching practices that put a premium on the imaginative aspects of learning, if we encouraged children to maintain that wonderful fantasy and speculative ability they possessed when they started school, we would be better able to create a culture that was much more receptive to the possibilities of human experience and much more educationally productive.

In speaking of the development of the imagination and the enhancement of wonder, I risk the possibility of seeming to talk about aims that are "pie in the sky." I do not believe such aims are "pie in the sky." I believe they can be attained. Their attainment will require a shift in what we consider really important in school. Equally as important, we will need to invent curriculum activities that afford youngsters opportunities to use their imagination and to engage in wonder. I have no doubt whatsoever that such activities can be designed, even in subjects that appear to be literal and highly rule-governed.

### *The School as Community*

A fifth aim of fundamental importance in schools is helping children realize that they are part of a caring community. Such a realization will not occur unless the school itself becomes a community. Many schools lack a sense of community; and many students, particularly during the time they are most vulnerable, their adolescence, are in institutions in which sustained and intimate contact with a caring adult is limited. Departmentalized school structures, which are the norm at the middle and senior high school levels, often provide adolescents with no adults in a counseling capacity nor any who are responsible for their pastoral care. Typically, a student will have five to seven teachers, will often see these teachers each day for 45 to 50 minutes, and in some schools will have access to a school counselor, but more often than not when such contact occurs, it is for purposes of vocational or academic counseling or for discipline. In few schools is there someone in touch with the emotional and social aspects of the students' lives. Whether we're students or teachers, all of us inevitably bring our personal lives into our classrooms.

The absence of pastoral care for students might not be so significant if the communities in which they lived were characterized by strong nuclear families and strong social bonds. The fact of the matter is that 25 percent of school-aged children are raised in single-parent homes; 52 percent of minority school children are raised in single-parent homes. Very often these children come home to a home with no one at home. We call them latchkey children. The seven to eight million latchkey children in America are twice as likely to use drugs as those who have a parent working at home. Since we have approximately 46 million students in public schools, seven to eight million latchkey children represents approximately 20 percent of our school population. In short, the school has always had a caring role to play, but this role is considerably more critical today. Developing an ethic of caring and creating a community that cares is, as far as I know, on no one's list of educational priorities—but it ought to be.

**My hope is that educators will be moved to begin the kind of dialogue that leads to genuine reform in education, a reform that pays attention to what really counts.**

### *Each Student's Personal Signature*

A sixth aim I believe to be of fundamental importance in schools is teaching children that they have a unique and important personal signature. Much that we teach by virtue of the curriculum content we emphasize and how we choose to evaluate what students have learned diminishes their sense of personal signature. The tasks we use to teach beginning arithmetic, spelling, punctuation, writing, and even reading emphasize the acquisition of social conventions. It is, of course, appropriate for schools to assign great importance to enabling the young to learn those social conventions we call the three R's. One of the features of these social conventions is that the child's task is to take conventions from the outside and, so to speak, to internalize them. These tasks can be regarded as requiring the skills of *impression*: they are intended to impress the child. And indeed they must. There are no teachers I know who seek from their students ingenuity in spelling. We want students to be able to communicate, to possess a shared set of conventions, and knowing which letters are to appear in what words in what order is a part of that aspiration. And it should be.

The problem is that such tasks dominate school programs. Children need opportunities not only to acquire the skills of impression, but also the skills of *expression*. In expressive activities the source of content is not located primarily outside the child. It is not simply a matter of acquiring those skills that demonstrate that social conventions have been learned but, rather, helping students reach down into their unique beings in order to find content that can be made visible in the public world. Again, the fine arts, including creative writing, are fields in which personal signature is particularly important. When a teacher gives a class of students the same words to spell, the pedagogical end-in-view is uniformity of outcome. The teacher wants all of the students to spell all of the words correctly, and that means the same way.

When that same teacher asks children to write a creative story, or paint a landscape, or choreograph a dance, or compose a piece of music, the last thing the teacher wants is uniformity of outcome. What the teacher seeks is heterogeneity, diversity, idiosyncracy, works that attest to the distinctive ways in which individual children see, feel, and imagine.

I am not arguing that such tasks should monopolize our programs, but we should at least have some semblance of balance. The provision of such opportunities and the inclusion of the fine arts and related subjects make it possible for children not only to realize, some for the first time, their own positive uniqueness, but also to find a place in our educational sun. Equity is achieved in education not only by giving students an opportunity to come to school, it is also influenced by what they find when they arrive. School programs that create a very narrow eye of a needle through which all children must pass diminish educational equity. Thus, the social—and indeed the moral conditions—that ought to prevail in our schools are those that broaden the eye of the needle and make it possible for all children to discover their aptitudes. Such programs would expand considerably our conception of giftedness. Indeed, if we provided opportunities in school for them to display their interests and talents, we would find that virtually all children were gifted in some way. At present, the game is far too narrow.

Related to this outcome is another, and that is the view that the genuinely good school does not diminish individual differences, it expands them. Virtue in the context of education is not achieved by bringing all children to the same destination; it is achieved by helping them learn how to become who they are. Variance, not homogeneity, is what counts in schooling, and we ought to design the environments and invent the activities that will make that variance possible.

## Our Legacy to Students

The aims that I have identified concerning what really counts in schools are not those inscribed in *A Nation at Risk* (NCEE 1983), they are not on the list generated by President Bush during his September 1989 education summit, and, as far as I know, they are not on any state department of education's educational priorities. Yet, for me, they are central to any adequate conception of education. The realization of such aims will require more than the appearance of articles in *Educational Leadership* or speeches at ASCD's Annual Conference. The centerpiece of change is the teacher and the central location, the classroom and the school.

I have no doubt that the creation of the conditions instrumental to the achievement of these values is possible. I also have no doubt that their realization will require the active engagement of teachers and the serious support of school administrators. It will, most certainly, also require the support of communities, who often need to understand better than they do at present what education is about and what really counts in schools, not only in the short term, but in the long run.

My hope is that educators will be moved to begin the kind of dialogue that leads to genuine reform in education, a reform that pays attention to what really counts. Genuine reform will require attention not only to the mission of our schools, but to their educational ecology. It will require attention not only to our intentions, but to the ways in which we organize our workplaces, to the scope of the programs that we provide, to the quality of our teaching, and to the means through which we assess what really matters. The agenda is large, formidable, and important. But the creation of that agenda and its realization are the legacy that we have the privilege to leave to our students. What other legacy could be more important?☐

*References*

Apple, M. (1982). *Education and Power.* Boston: Routledge and Kegan Paul.

Arendt, H. (1958). *The Human Condition.* New York: Doubleday.

Dreeben, R. (1968). *On What Is Learned in School.* New York: Addison-Wesley.

Eisner, E. (1985). *The Educational Imagination.* 2nd ed. New York: Macmillan.

National Commission on Excellence in Education. (1983). *A Nation at Risk.* Washington, D.C.: U.S. Government Printing Office.

**Elliot W. Eisner** is Professor of Education and Art, Stanford University, School of Education, Stanford, CA 94305.

# III-5

# Caring Kids: The Role of the Schools

Alfie Kohn
Belmont, Massachusetts

First published in *Phi Delta Kappan* 72, 7 (March 1991): 496–506. Copyright © 1991 by Alfie Kohn. Reprinted with the author's permission.

# Caring Kids

## The Role of the Schools

*Psychological research, common sense, and the experience of an important pilot project in California offer specific guidance for helping children to grow into caring adults. Mr. Kohn provides the details.*

BY ALFIE KOHN

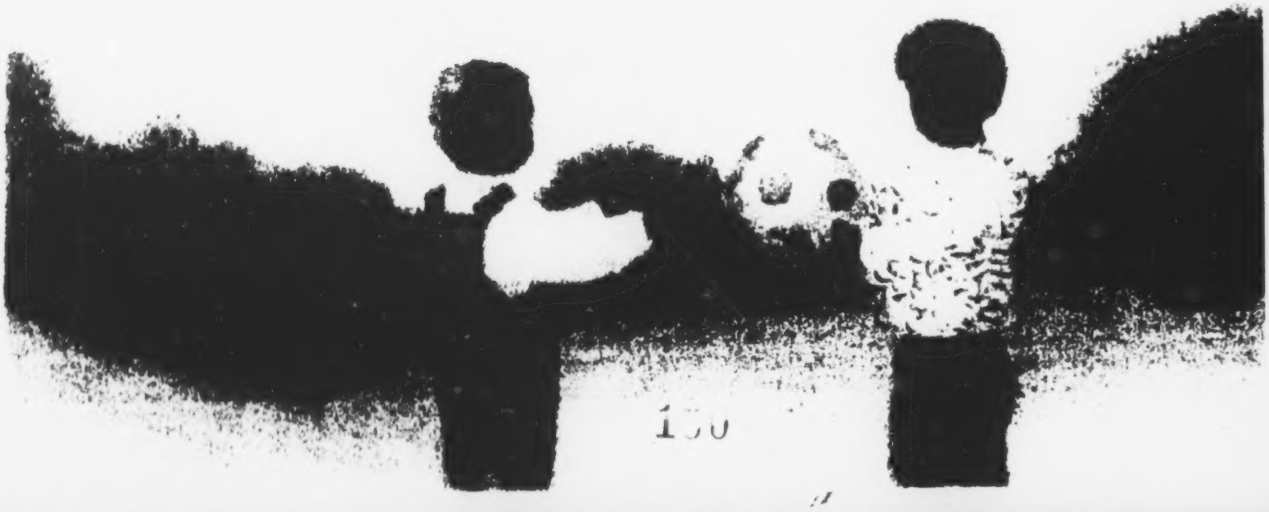
*Illustration by Rob Colvin*

EDUCATION worthy of the name is essentially education of character," the philosopher Martin Buber told a gathering of teachers in 1939.[1] In saying this, he presented a challenge more radical and unsettling than his audience may have realized. He did not mean that schools should develop a unit on values or moral reasoning and glue it onto the existing curriculum. He did not mean that problem children should be taught how to behave. He meant that the very profession of teaching calls on us to try to produce not merely good learners but good people.

Given that even the more modest task of producing good learners seems impossibly ambitious – perhaps because of a misplaced emphasis on producing good test-takers – the prospect of taking Buber seriously may seem posi-

---

*ALFIE KOHN is an independent scholar living in Cambridge, Mass., who writes and lectures widely on human behavior and education. His books include* The Brighter Side of Human Nature: Altruism and Empathy in Everyday Life *(Basic Books, 1990) and* No Contest: The Case Against Competition *(Houghton Mifflin, 1986). ©1991, Alfie Kohn.*



The belief persists in this culture that our darker side is more pervasive . . .

tively utopian. But in the half-century since his speech, the need for schools to play an active role in shaping character has only grown more pressing. That need is reflected not only in the much-cited prevalence of teenage pregnancy and drug use but also in the evidence of rampant selfishness and competitiveness among young people.* At a tender age, children learn not to be tender. A dozen years of schooling often do nothing to promote generosity or a commitment to the welfare of others. To the contrary, students are graduated who think that being smart means looking out for number one.

I want to argue, first, that something *can* be done to rectify this situation because nothing about "human nature" makes selfishness inevitable; second, that educators in particular *should* do something about the problem; and third, that psychological research, common sense, and the experience of an important pilot project in California offer specific guidance for helping children to grow into caring adults.

MUCH OF what takes place in a classroom, including that which we have come to take for granted, emerges from a set of assumptions about the nature of human nature. Not only how children are disciplined, but the very fact that influencing their actions is viewed as "discipline" in the first place; not merely how we grade students, but the fact that we grade them at all; not simply how teachers and students interact, but the fact that interaction *between students* is rarely seen as integral to the process of learning – all of these facts ultimately rest on an implicit theory of what human beings are like.

Consider the fact that most conversations about changing the way children act in a classroom tend to focus on curbing negative behaviors rather than on promoting positive ones. In part, this emphasis simply reflects the urgency of preventing troublesome or even violent conduct. But this way of framing the issue may also tell us something about our view of what comes naturally to children, what they are capable of, and, by extension, what lies at the core of our species. Likewise, it is no coincidence, I think, that the phrase "it's just human nature to be . . ." is invariably followed by such adjectives as selfish, competitive, lazy, aggressive, and so on. Very rarely do we hear someone protest, "Well, of course he was helpful. After all, it's just human nature to be generous."

The belief persists in this culture that our darker side is more pervasive, more persistent, and somehow more real than our capacity for what psychologists call "prosocial behavior." We seem to assume that people are naturally and primarily selfish and will act otherwise only if they are coerced into doing so and carefully monitored. The logical conclusion of this world view is the assumption that generous and responsible behavior must be forced down the throats of children who would otherwise be inclined to care only about themselves.

A review of several hundred studies has convinced me that this cynicism is not realism. Human beings are not only selfish and self-centered, but also decent, able to feel – and prepared to try to relieve – the pain of others. I believe that it is as "natural" to help as it is to hurt, that concern for the well-being of others often cannot be reduced to self-interest, that social structures predicated on human selfishness have no claim to inevitability – or even prudence. This is not the place for rehearsing the arguments and data that support these conclusions – in part because I have recently done so at book length.[2] But I would like to mention a few recent findings from developmental psychology that speak to the question of whether educators can aim higher than producing a quiet classroom or a nondisruptive child.

To start at the beginning, newborns are more likely to cry – and to cry longer – when they are exposed to the sound of another infant's crying than when they hear other noises that are equally loud and sudden. In three sets of studies with infants ranging in age from 18 to 72 hours, such crying seemed to be a spontaneous reaction rather than a mere vocal imitation.[3] In the view of Abraham Sagi and Martin Hoffman, who conducted one of the studies, this finding suggests the existence of "a rudimentary empathic distress reaction at birth."[4] Our species may be primed, in other words, to be discomfited by someone else's discomfort.

As an infant grows, this discomfort continues and takes more sophisticated forms. Marian Radke-Yarrow, Carolyn Zahn-Waxler, and their associates at the National Institute of Mental Health have been studying toddlers for nearly 20 years, having in effect deputized mothers as research assistants to collect data in the home instead of relying on brief (and possibly unrepresentative) observations in the laboratory. A 10- to 14-month-old child, they have found, can be expected to show signs of agitation and unhappiness in the presence of another person's distress, perhaps by crying or burying her head in her mother's lap. As a child develops the capacity to undertake more deliberate behavior, in the period between 18 and 24 months, his response to distress will become more active: patting the head, fetching a toy, offering verbal expressions of sympathy, finding an adult to help, and so forth.[5]

I should add that, like all parents, researchers have also observed hostile and selfish actions on the part of children. To say that sympathy or helping behavior is pervasive and precocious is not to claim that every child is an angel or to deny that toddlers – particularly in a society preoccupied with possessiveness – will sometimes snatch back a toy ("Mine!") or throw it across the room. But it is to argue that the *antisocial* is no more basic or natural than the *prosocial*.

By the time children are of preschool age, comforting, sharing, and helping are regular occurrences. One study of pre-

---

*Our society's current infatuation with the word *competitiveness*, which has leached into discussions about education, only exacerbates the problem by encouraging a confusion between two very different ideas: excellence and the desperate quest to triumph over other people.

schoolers during free play discovered that 67 of the 77 children shared with, helped, or comforted another child at least once during only 40 minutes of observation.[6] After counting such behaviors in similar experiments of her own, Arizona State University psychologist Nancy Eisenberg became curious about why children were acting this way. To find out, she came up with a technique that few research psychologists had thought to use: she *asked* the children.

Eisenberg and a colleague simply followed 4- and 5-year-olds around a preschool and watched for unprompted prosocial behavior. Each time such an act was observed, the child was asked why he or she did it. ("How come you gave that to John?") None of the children intended to conform to adult expectations or expressed any fear of punishment. Very few said that they expected to benefit in some way by helping – such as by impressing their peers. Among the most frequent explanations heard was the simple observation that the other child had needed help.[7] This, when you come right down to it, is the heart of altruism.[8] And it is enough to suggest that parents and educators hoping to raise a child who is responsive to the needs of others already have "an ally within the child," in Martin Hoffman's lovely phrase.

There is nothing about caring for others that implies not caring for oneself.

IF WE HAD to pick a logical setting in which to guide children toward caring about, empathizing with, and helping other people, it would be a place where they would regularly come into contact with their peers and where some sort of learning is already taking place. The school is such an obvious choice that one wonders how it could be that the active encouragement of prosocial values and behavior – apart from occasional exhortations to be polite – plays no part in the vast majority of American classrooms. This would seem to stem either from a lack of interest in the idea or from some objection to using the schools in particular for this purpose. Both factors probably play a role, but I will concentrate here on the latter and consider three specific reservations that parents, teachers, policy makers, and others may have – or at least may hear and thus need to answer – about classroom-based programs to help children develop a prosocial orientation.

The first objection is that an agenda concerned with social and moral issues amounts to teaching values – a dangerous business for a public institution. In response, we must concede that a prosocial agenda is indeed value-laden, but we should immediately add that the very same is true of the status quo. The teacher's presence and behavior, her choice of text, the order in which she presents ideas, and her tone of voice are as much part of the lesson as the curriculum itself. So, too, is a teacher's method of discipline or classroom management saturated in values, regardless of whether those values are transparent to the teacher. In short, to arrange our schools so that caring, sharing, helping, and empathizing are actively encouraged is not to introduce values into a neutral environment; it is to examine the values already in place and to consider trading them in for a new set.

It is sometimes said that moral concerns and social skills ought to be taught at home. I know of no one in the field of education or child development who disagrees. The problem is that such instruction – along with nurturance and warmth, someone to model altruism, opportunities to practice caring for others, and so forth – is not to be found in all homes. The school may need to provide what some children will not otherwise get. In any case, there is no conceivable danger in providing these values in both environments. Encouragement from more than one source to develop empathic relationships is a highly desirable form of redundancy.

The second concern one hears – and this one dovetails with the broader absence of interest in the prosocial realm – is the fear that children taught to care about others will be unable to look out for themselves when they are released into a heartless society. The idea that someone exposed to such a program will grow up gullible and spineless, destined to be victimized by mean-spirited individuals, can be traced back to the prejudice that selfishness and competitiveness are efficacious social strategies – a sterling example of what sociologist C. Wright Mills used to call "crackpot realism." In fact, those whose mantra is "look out for number one" are actually at a greater disadvantage in any sort of society than those who are skilled at working with others and inclined to do so. Competition and the single-minded pursuit of narrowly conceived self-interest typically turn out to be counterproductive.

By contrast, a well-designed program of prosocial instruction will include training in cooperative conflict resolution and in methods of achieving one's goals that do not require the use of force or manipulation. But even without such a component, there is nothing about caring for others that implies not caring for or looking after oneself. A raft of research has established that assertiveness, healthy self-esteem, and popularity are all compatible with – and often even correlates of – a prosocial orientation.[9]

The final objection to teaching children to be caring individuals is that the time required to do so comes at the expense of attention to academics – a shift in priorities apt to be particularly unpopular at a time when we entertain ourselves by describing how much students don't know. In fact, though, there is absolutely no evidence to suggest that prosocial children – or the sort of learning experiences that help to create them – are mutually exclusive with academic achievement. To the contrary, the development of perspective-taking – the capacity to imagine how someone else thinks, feels, or sees the world – tends to promote cognitive problem solving generally. In one study, the extent to which girls had these skills at age 8 or 9 was a powerful predictor of performance on reading and spelling tests taken two years later – an even better predictor, in fact, than their original test scores.[10]

Not only are the ingredients of a prosocial orientation conducive to academic excellence, but the educational process itself does not require us to choose between teaching children to think and teaching them to care. It is possible to integrate prosocial lessons into the regular curriculum; as long as children are learning to read and spell and think critically, they may as well learn with texts that encourage perspective-taking. Indeed, to study literature or history by grappling with social and moral dilemmas is to invite a deeper engagement with those subjects. Meanwhile, literally hundreds of studies have shown that cooperative learning, which has an important place in a prosocial classroom, enhances achievement regardless of subject matter or age level.[11] So consistent and remarkable have these results been that schools and individual teachers often adopt models of cooperative learning primarily to strengthen academic performance. The development of prosocial values is realized as an unintended bonus.

Like sticks, carrots are artificial attempts to manipulate behavior.

EDUCATION of character in Buber's sense asks of teachers something more than the mere elimination of behavior problems in the classroom. The absence of such problems is often seen as an invitation to move past behavioral and social issues and get on with the business at hand, which is academic learning. I am arguing, by contrast, that behavioral and social issues, values and character, are very much part of the business at hand. But whether we are talking about addressing misconduct or about taking the initiative to help students become more responsive to one another, a teacher can take any of several basic orientations. Here are four approaches to changing behaviors and attitudes, presented in ascending order of desirability.

**1. Punishing.** A reliance on the threat of punishment is a reasonably good indication that something is wrong in a classroom, since children have to be bullied into acting the way the teacher demands. Apart from the disagreeable nature of this style of interaction – which cannot be disguised, incidentally, by referring to punishment as "consequences" – it is an approach distinguished mostly by its ineffectiveness. Decades of research have established that children subjected to punitive discipline at home are *more* likely than their peers to break rules when they are away from home.

Isolating a child from his peers, humiliating her, giving him an F, loading her with extra homework, or even threatening to do any of these things can produce compliance in the short run. Over the long haul, however, this strategy is unproductive.

Why? First, at best, punishment teaches nothing about what one is supposed to do – only about what one is not supposed to do. There is an enormous difference between not beating up one's peers, on the one hand, and being helpful, on the other.

Second, the child's attention is not really focused on the intended lesson at all ("pushing people is bad"), much less on the rationale for this principle, but primarily on the punishment itself. Figuring out how to get away with the misbehavior, how to avoid detection by an authority, is a perfectly logical response. (Notice that the one who punishes becomes transformed in the child's eyes into a rule-enforcer who is best avoided.) Social learning theory tells us that this attention to the punishment is also likely to *teach* the child to be punitive and thus exacerbate the behavior problems; a teacher's actions do indeed speak louder than words.

Finally, punishment breeds resistance and resentment. "The more you use power to try to control people, the less real influence you'll have on their lives," Thomas Gordon has written.[12] Since such influence is associated with helping children to develop good values, the use of power would seem ill-advised.

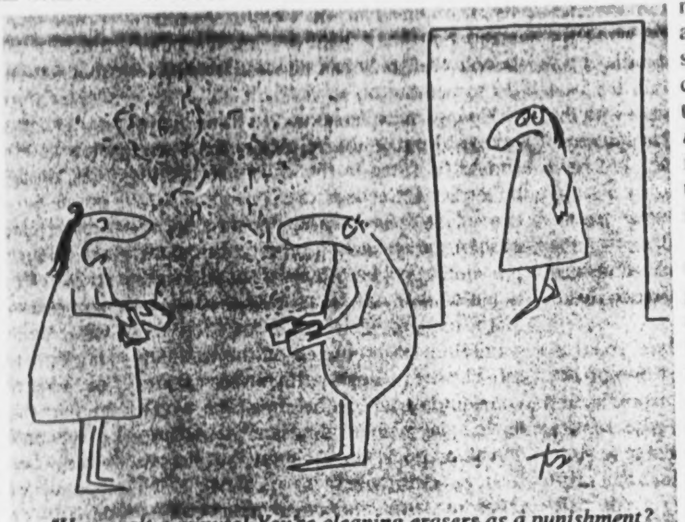
**2. Bribing.** There is no question that rewards are better than punishment. On the other hand, what these two methods share is probably more important than the respects in which they differ, and herein lies a tale that will be highly disconcerting to educators enamored of positive reinforcement. Psychological – and particularly developmental – theory and research have come a long way since the simplistic behaviorism of the last generation, but many well-meaning teachers continue to assume that what works for training the family pet must be appropriate for shaping children's actions and values as well.

Gold stars, smiley faces, trophies, certificates, high grades, extra recess time, candy, money, and even praise all share the feature of being "extrinsic" to whatever behavior is being rewarded. Like sticks, carrots are artificial attempts to manipulate behavior that offer children *no reason to continue acting in the desired way when there is no longer any goody to be gained*. Do rewards motivate students? Absolutely. They motivate students to get rewarded. What they fail to do is help children develop a commitment to being generous or respectful.

In fact, the news is even worse than this. Not only is bribing someone to act in a particular way ultimately ineffective, but, like the use of threats, it can actually make things worse. Consider the effects of rewards on achievement. Yale psychologist Robert Sternberg recently summed up what a growing number of motivation researchers now concede: "Nothing tends to undermine creativity quite like extrinsic motivators do. They also undermine intrinsic motivation: when you give extrinsic rewards for certain kinds of behavior, you tend to reduce children's interest in performing those behaviors for their own sake."[13] Once we see ourselves as doing something in order to get a reward, we are less likely to want to continue doing it in the absence of a reward – even if we used to find it enjoyable.

Readers of the *Kappan* were first exposed to research demonstrating this phenomenon more than 15 years ago,[14] and the data have continued to accumulate since then, with some studies concentrating on how extrinsic motivators reduce intrinsic interest and others showing how they undermine performance, particularly on creative tasks.[15] A number of explanations have been proposed to account for these remarkably consistent findings. First, people who think of themselves as working for a reward feel controlled, and this lack of self-determination interferes with creativity. Second, rewards encourage "ego involvement" to the exclusion of "task involvement," and the latter is more predictive of achievement. Third, the promise of a reward is "tantamount to declaring that the activity is not worth doing for its own sake," as A. S. Neill put it;[16] indeed, anything construed as a prerequisite to some other goal will probably be devalued as a result.

What is true for academic learning also applies to behavior. A little-known series of studies has pointed up the folly of trying to encourage prosocial behavior through the use of extrinsic incentives. Children who received rewards for donating to another child – and, in another experiment, adults who were paid for helping a researcher – turned out to be less likely to describe themselves in words suggesting intrinsic motivation to help than were people who received nothing in return.[17] In another study, women offered money for answering a questionnaire were less likely to agree to a similar request two or three days later, when no money was involved, than were women who had not been paid for helping with the first survey.[18]



***"Hey, wait a minute! You're cleaning erasers as a punishment? I'm cleaning erasers as a reward!"***

The implication is that, when someone is rewarded for prosocial behavior, that person will tend to assume that the reward accounts for his or her actions and thus will be less likely to help once no one is around to hand out privileges or praise. Indeed, elementary school students whose mothers believed in using rewards to motivate them were less cooperative and generous than other children in a recent study.[19] Such findings are of more than theoretical interest given the popularity of Skinnerian techniques for promoting generosity in schools. A recent *New York Times* article described elementary schools where helpful children have their pictures posted in hallways, get to eat at a special table in the cafeteria, or even receive money.[20] Such contrivances may actually have the effect of undermining the very prosocial orientation that their designers hope to promote.

**3. Encouraging commitment to values.** To describe the limitations of the use of punishments and rewards is already to suggest a better way: the teacher's goal should not be simply to produce a given behavior – for example, to get a child to share a cookie or stop yelling – but to help that child see himself or herself as the kind of person who is responsible and caring. From this shift in self-concept will come lasting behaviors and values that are not contingent on the presence of someone to dispense threats or bribes. The child has made these behaviors and values his or her own.

A student manipulated by currently fashionable behavioral techniques, however, is unlikely to internalize the values underlying the desired behaviors. At the heart of "assertive discipline," for example, is control: "I want teachers to learn that they have to take charge," Lee Canter explained recently.[21] I don't. I want *children* to become responsible for what they do and for what kind of people they are. The teacher has a critical role to play in making sure that this happens; in criticizing manipulative approaches I am not suggesting that children be left alone to teach themselves responsibility. But the teacher ought to be guided less by the need to maintain control over the classroom than by the long-term objective of helping students to act responsibly because they understand that it is right to do so.

I will have more to say below about strategies for facilitating this internalization, but first I want to mention a version of this process that I believe is even more desirable – the ideal approach to helping children become good people.

**4. Encouraging the group's commitment to values.** What the first two approaches have in common is that they provide nothing more than extrinsic motivation. What the first two share with the third is that they address only the individual child. I propose that helpfulness and responsibility ought not to be taught in a vacuum but in the context of a community of people who learn and play and make decisions together. More precisely, the idea is not just to internalize good values *in* a community but to internalize, among other things, the value *of* community.

Perhaps the best way to crystallize what distinguishes each of these four approaches is to imagine the question that a child is encouraged to ask by each. An education based on punishment prompts the query, "What am I supposed to do, and what will happen to me if I don't do it?" An education based on rewards leads the child to

> **From preschool to high school, children should learn why helping others is good.**

ask, "What am I supposed to do, and what will I get for doing it?" When values have been internalized by the child, the question becomes "What kind of person do I want to be?" And, in the last instance, the child wonders: "How do we want our classroom (or school) to be?"

EDUCATORS eager to have children think about how they want their classrooms to be – which is to say, educators who do not feel threatened at the prospect of inviting children to share some of the responsibility for creating norms and determining goals – need to think in terms of five broad categories: what they believe, what they say, what they do, how they relate to students, and how they encourage students to relate to one another. Let us consider each in turn.

*What educators believe.* The famous Pygmalion effect refers to the fact that a teacher's assumptions about a child's intellectual potential can affect that child's performance. Such self-fulfilling prophecies, however, are by no means limited to academics; they also operate powerfully on a child's actions and values. Write off a student as destructive or disruptive, and he or she is likely to "live down to" these expectations. Conversely – and here is the decisive point for anyone concerned with promoting generosity – attributing to a child the best possible motive that is consistent with the facts may set in motion an "auspicious" (rather than a vicious) circle. We help students develop good values by assuming whenever possible that they are already motivated by these values – rather than by explaining an ambiguous action in terms of a sinister desire to make trouble.

However, what we assume about a given student is also colored by our assumptions regarding human nature itself. While I am not aware of any research on this question, it seems reasonable to suppose that an educator who thinks that self-interest motivates everything we do will be suspicious of individual instances of generosity. Someone who takes for granted that a Hobbesian state of nature would exist in a classroom in the absence of a controlling adult to keep children in line, who believes that children need to be leaned on or "taught a lesson" or bribed to act responsibly, is likely to transfer these expectations to the individual child and to produce an environment that fulfills them. The belief that children are actually quite anxious to please adults, that they may simply lack the skills to get what they need, that they will generally respond to a caring environment can create a very different reality. What you believe matters.

*What educators say.* An immense body of research has shown that children are more likely to follow a rule if its rationale has been explained to them and that, in general, discipline based on reason is more effective than the totalitarian approach captured by the T-shirt slogan "Because I'm the mommy, that's why." This finding applies not only to discouraging aggression but to promoting altruism. From preschool to high school, children should learn why – not merely be told that – helping others is good. Pointing out how their actions affect others sensitizes students to the needs and feelings of others and tacitly communicates a message of trust and responsibility. It implies that, once children understand how their behavior makes other people feel, they can and will choose to do something about it.

How such explanations are framed also counts. First, the level of the discourse should be fitted to the child's ability to understand. Second, the concept of using reason does not preclude passion. A prohibition on hurting people, for example, should not be offered dispassionately but with an emotional charge to show that it matters. Third, prosocial activity should not be promoted on the basis of self-interest. "Zachary, if you don't share your dump truck with Linda, she won't let you play with her dinosaur" has an undeniable appeal for a parent, but it is a strategy more likely to inculcate self-regarding shrewdness than genuine concern for others. The same goes for classroom exhortations and instruction.

A series of studies by Joan Grusec of the University of Toronto and her colleagues is also relevant. Her research provides a concrete alternative to the use of rewards or praise to elicit generosity. "Children who view their prosocial conduct as compliance with external authority will act prosocially only when they believe external pressures are present," she has written. Far preferable is for children to "come to believe that their prosocial behavior reflects values or dispositions in themselves."[22]

This result is best achieved by verbally attributing such values or dispositions to the child. In one experiment, in which children gave away some of their game winnings after watching a model do so, those who were told that they had made the donation "because you're the kind of person who likes to help other people" were subsequently more generous than those who were told that they had donated because they were expected to do so.[23] In another study, the likelihood of children's donating increased both when they were praised and when they were led to think of themselves as helpful people. But in a follow-up experiment, it was the latter group who turned out to be more generous than those who had received verbal reinforcement. In other words, praise increased generosity in a given setting but ceased to be effective outside of that setting, whereas children with an intrinsic impulse to be generous continued to act on that motivation in other circumstances.[24]

A study of adults drives home the point. Subjects who were told that a personality test showed that they were kind and thoughtful people were more likely to help a confederate who "accidentally" dropped a pile of cards than were those who were told that they were unusually intelligent or those given no feedback at all. This finding is important because it implies that being led to think of oneself as generous does not affect behavior merely because it is a kind of reinforcement or a mood-enhancer; this label apparently encourages prosocial action because it helps to build a view of the self as altruistic.[25]

This is not to suggest that a teacher's every utterance must be – or can be – geared toward internalization. Simply making sure that a classroom is a safe environment conducive to learning can require the sort of behavioral interventions on a day-to-day basis that don't do much to strengthen a child's prosocial self-concept. But the more teachers attend to the latter, the fewer problems they are likely to have over the long run.

*What educators do.* Children of all ages, from before the time they can read until after the time they start seeking distance from adults, learn from what they see. Studies show that children who watched, even briefly, as someone donated to charity were themselves likely to donate more than other children – even if months had elapsed since the exposure to the model.[26] The extent to which a teacher expresses concern about people in distress and takes the initiative to help – which applies both to how the teacher treats the students themselves and how he or she refers to people outside the classroom – can set a powerful example and be even more effective than didactic instruction in promoting a sense of caring in students.

There is no shortage of suggestions about how to devise lessons that address social and ethical issues, ranging from explicit training in perspective-taking or moral reasoning to discussions about values that can, in turn, include either "clarification" of the beliefs that students already hold or old-fashioned lectures on character or morality. Most of the debate on the subject occurs between proponents of just such programs, each accusing the other of being relativistic or of seeking to indoctrinate. Far less consideration is given to the possibility of integrating such issues into the regular curriculum.

A distinction, though not a sharp one, can be made between teaching morality (or about morality) as such and helping children to be positively connected to others. The latter is my focus here, and some writers have argued that, particularly for younger children, it ought to be the primary focus in the schools, too. "Unless the young child has acquired a positive propensity towards other persons," says one educator, "subsequent moral education will become virtually impotent."[27]

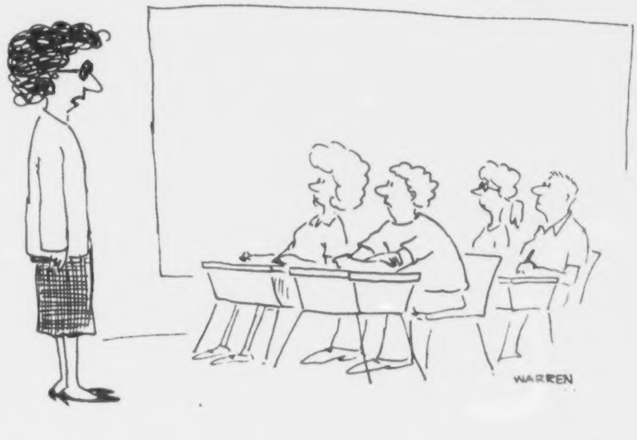
As an alternative to special units devoted to one of these approaches, children can use texts in conventional subject areas that encourage perspective-taking. This option should allay the concern that moral instruction will distract us from academics.

*How educators relate to students.* Preceding and underlying specific techniques for encouraging particular behaviors is the practice of nesting all kinds of discipline and instruction in the context of a warm, nurturant, and empathic relationship with students. Children whose parents are interested in and supportive of them usually distinguish themselves as socially competent and psychologically healthy on a range of measures, and there is no reason to think that the teacher/student relationship is any different.

Warm, caring, empathic adults do several things at once. They provide the child with a benevolent, safe place in which to act. (If a child's experience with others leaves him or her feeling threatened rather than safe, this is likely to foster psychological damage control at the expense of any inclination to help others.) I hope that few educators take seriously the absurd dictum that teachers should display no warmth until well into the school year – after firm control of the classroom has been won. Instead, teachers should establish themselves from the beginning as the students' allies, adults with whom they can work to solve the problems that emerge during the normal course of development. In meeting a child's emotional needs we give him or her the emotional freedom to meet the needs of others.

*How educators encourage students to relate to one another.* Anyone interested in children as social beings must recognize the need to attend to the interactions among them in the classroom. In most American schools, children are forced to work either against one another (by competing) or apart from one another (by learning individually). The chance to work *with* one another, to learn social skills and caring, is left to happen by itself during recess, at lunch, or after school. This single fact goes a long way toward explaining why people in our society tend to regard others as potential obstacles to their own success. David Johnson and Roger Johnson of the University of Minnesota have emphasized that the relationship between student and student is at least as important as that between student and teacher or between student and curriculum. Their reference to student/student relationships is offered in the context of creating good learners, and it is all the more true in terms of creating good human beings.

How, specifically, should teachers encourage student interaction? First, students can be put in pairs or small groups so as to help one another learn. This concept, known as cooperative learning, embraces many disparate models for implementation: some depend on grades and other extrinsic incentives to insure that students work together, some involve cooperation among as well as within groups, some provide for a strict division of labor in completing assignments. A substantial number of studies have found that cooperative learning of various types



*"This class will stimulate your ideas and thoughts. And remember – no talking."*

**Do quiet children learn more effectively or merely make fewer demands on teachers?**

has the potential to help students feel good about themselves, feel good about each other, feel good about what they are learning, and learn more effectively.

Cooperation, by virtue of being an interaction in which two or more people work together for mutual benefit, is not itself an example of prosocial behavior as the term is usually used. Neither does its successful use presuppose the existence of prosocial motives in all children. Rather, by creating interdependence and a built-in incentive to help, cooperative learning *promotes* prosocial behavior. Having children learn from one another creates powerful bonds between them and sends a message very different from that sent by a classroom in which each child is on his or her own – or, worse still, one in which the success of each is inversely related to the success of the others.[28]

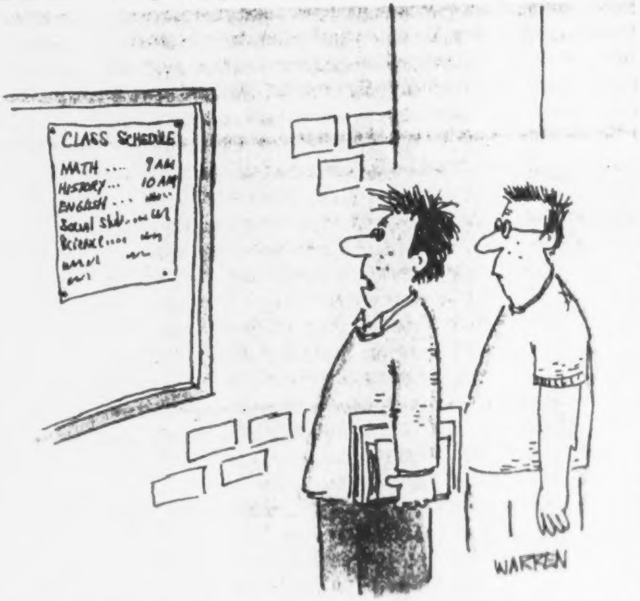
In one study, fifth-graders who studied grammar in cooperative learning groups were more likely to give away prize tokens to a stranger than were those who studied on their own; in another, kindergartners who participated in cooperative activities acted more prosocially than their peers in a traditional classroom.[29] But the consequences are not limited to generosity per se. Carefully structured cooperative learning also promotes a subjective sense of group identity, a greater acceptance of people who are different from oneself (in terms of ethnicity or ability level), and a more sophisticated ability to imagine other people's points of view.[30] Cooperation is an essentially humanizing experience that predisposes participants to take a benevolent view of others. It allows them to transcend egocentric and objectifying postures and encourages trust, sensitivity, open communication, and prosocial activity.

Second, teachers can move the idea of discipline not only away from punishments and rewards but also away from the premise of these strategies – namely, that teachers should simply be figuring out by themselves how to elicit compliance with a set of rules or goals that they alone devise. The realistic alternative is not for the teacher to abdicate responsibility for what happens in the classroom but rather to bring in (and guide) children so that they can play a role in making decisions about how their classroom is to be run and why. (Must hands always be raised or only during certain kinds of discussions? What is the best way for the class as a community to balance principles of fairness and the spontaneity that encourages participation?)

Discipline would thus be reconfigured as collaborative planning and mutual problem solving. Such an approach will be preferred by anyone who favors the idea of autonomy and democratic decision making – but it can also be argued that purely practical considerations recommend it, since children are more likely to follow rules that they have helped to create than rules dictated to them. This, of course, assumes that following rules is in itself a desirable goal. More broadly, educators need to ask themselves and each other about the ultimate objective of discipline. Even if one of the conventional programs of behavior control succeeded in keeping children quiet, do quiet children learn more effectively or merely make fewer demands on the teacher? (The Johnsons like to say that a principal walking through the school corridors should be concerned if he or she hears no sound coming from a classroom; this means that real learning probably is not taking place.) And which approach is most likely to help children come to care about one another?

To invite children to participate in making decisions not only about classroom procedures but also about pedagogical matters (what is to be learned, how, and why) and housekeeping matters (how to celebrate birthdays or decorate the walls) is to bring them into a process of discussion, an opportunity to cooperate and build consensus. To this extent, it is a chance for them to practice perspective-taking skills, to share and listen and help. In short, involving children in planning and decision making is a way of providing a framework for prosocial interactions that supports other such opportunities; it turns a routine issue into another chance to learn about and practice caring – and, not so incidentally, thinking as well.[31]

Finally, educators can provide students with opportunities to be responsible for one another so that they will learn (prosocial values and skills) by doing. Ideally, this can include interaction with those of different ages. For an older child to guide someone



*"The only thing that keeps me going is the dream that one day there'll be a classless society."*

BEST COPY AVAILABLE

younger is to experience firsthand what it is to be a helper and to be responsible for someone who is dependent on him or her. For the younger child, this cross-age interaction presents an opportunity to see a prosocial model who is not an adult.

ONE OF THE most exciting and innovative educational programs now in operation, the Child Development Project (CDP), is devoted specifically to helping children become more caring and responsible.[32] The experience of the CDP offers lessons in the systematic application of many of the ideas discussed here; indeed, I owe my formulation of some of these ideas to the work done by Eric Schaps, Marilyn Watson, and others involved with the project.

> Prosocial values come from a synthesis of adult inculcation and peer interaction.

The CDP is the first long-term, comprehensive, school-based project in prosocial education. After being invited a decade ago to work in the San Ramon Valley (California) Unified School District, about 30 miles east of San Francisco, the staff carefully matched two sets of three elementary schools in the district for size and socioeconomic status. A coin flip then determined which of these sets would receive the program and which would serve as the comparison group. The first teachers were trained before the start of the 1982-83 school year. Staff researchers focused on a group of children in the experimental schools (then in kindergarten and now in junior high school) to assess whether their attitudes, behavior, and achievement differed significantly from those of their counterparts in the comparison schools. In the fall of 1988, the program was introduced into two elementary schools in nearby Hayward, a district more ethnically diverse than the white, affluent suburbs in San Ramon Valley, and Schaps is now seeking funding to take the program to eight more sites around the country.

"How do we want our classroom to be?" is exactly the question that the CDP would have children ask. Rejecting punishment and rewards in favor of strategies geared toward internalization of prosocial norms and values, the CDP invites teachers and students to work together to turn their classrooms into caring communities. The primary components of the program intended to bring this about are these:

- a version of cooperative learning that does not rely on grades or other extrinsic motivators;
- the use of a literature-based reading program that stimulates discussion about values and offers examples of empathy and caring even as it develops verbal skills;
- an approach to classroom management in which the emphasis is on developing intrinsic motives to participate productively and prosocially, in which teachers are encouraged to develop warm relationships with the children, and in which periodic class meetings are held so that children can play an active role in planning, assessing progress, and solving problems; and
- a variety of other features, including pairing children of different ages to work together, setting up community service projects to develop responsibility, giving periodic homework assignments specifically designed to be done (and to foster communication) with parents, and holding schoolwide activities that may involve whole families.

In their writings, members of the CDP staff have distinguished their way of teaching values from the approaches of better-known models. Unlike certain kinds of character education, the CDP approach emphasizes helping students understand the reason for a given value rather than simply insisting that they accept it or behave in a certain way because they have been told to do so. Unlike purely child-centered approaches, however, the CDP is committed to the importance of adult socialization: the teacher's job is to teach, to guide, to enforce, to facilitate cooperation, to model behaviors – in short, to be much more than a passive bystander. Prosocial values come from a synthesis of adult inculcation and peer interaction, and these values – in contrast to the programs developed by some theorists in the area of moral reasoning – emphasize caring for others as well as applying principles of fairness.

Prior to the implementation of the CDP, students randomly selected from the three experimental and the three comparison schools proved to be similar not only demographically but also on a range of social attitudes, values, and skills. Once the program was implemented, however, structured interviews and observations turned up significant differences between students participating in the program and those in the comparison schools on some, though not all, measures.

Children taking part in the CDP engaged in a greater number of spontaneous prosocial behaviors in class, seemed better able to understand hypothetical conflict situations, and were more likely to take everyone's needs into account in dealing with such situations. They were more likely to believe that one has an obligation to speak up in a discussion even if one's position seems unlikely to prevail (which should answer those concerned about the assertiveness of caring children). While the CDP's emphasis has not required any sacrifice of conventional achievement (as measured by standardized test scores), neither has it given participants a consistent academic advantage over students in comparison schools. (In part, this finding may be due to a ceiling effect: students in the district already score in the top 10% of California schoolchildren, so there is not much room for improvement.) By the time the CDP group reached sixth grade, though, they were outscoring their counterparts in the comparison schools on a measure of higher-order reading comprehension (essays written about stories and poems).

It remains to be seen whether and in what ways the values and behaviors of children from schools using the CDP will

BEST COPY AVAILABLE

continue to distinguish them from those who attended comparison schools now that they are all in junior high school. But this pilot project provides real evidence for the larger point I am making here: it is both realistic and valuable to attend to what students learn in the classroom about getting along with their peers. Children can indeed be raised to work with, care for, and help one another. And schools must begin to play an integral role in that process.

---

1. Martin Buber, *Between Man and Man*, trans. Ronald Gregor Smith (New York: Macmillan, 1965), p. 104.

2. Alfie Kohn, *The Brighter Side of Human Nature: Altruism and Empathy in Everyday Life* (New York: Basic Books, 1990).

3. Marvin L. Simner, "Newborn's Response to the Cry of Another Infant," *Developmental Psychology*, vol. 5, 1971, pp. 136-50; Abraham Sagi and Martin L. Hoffman, "Empathic Distress in the Newborn," *Developmental Psychology*, vol. 12, 1976, pp. 175-76; and Grace B. Martin and Russell D. Clark III, "Distress Crying in Neonates: Species and Peer Specificity," *Developmental Psychology*, vol. 18, 1982, pp. 3-9.

4. Sagi and Hoffman, p. 176.

5. See, for example, Carolyn Zahn-Waxler and Marian Radke-Yarrow, "The Development of Altruism: Alternative Research Strategies," in Nancy Eisenberg-Berg, ed., *The Development of Prosocial Behavior* (New York: Academic Press, 1982).

6. Marian Radke-Yarrow and Carolyn Zahn-Waxler, "Dimensions and Correlates of Prosocial Behavior in Young Children," *Child Development*, vol. 47, 1976, pp. 118-25.

7. Nancy Eisenberg-Berg and Cynthia Neal, "Children's Moral Reasoning About Their Own Spontaneous Prosocial Behavior," *Developmental Psychology*, vol. 15, 1979, pp. 228-29. Eisenberg and another colleague have observed that appeals to authority or punishment (which were completely absent here) are what one would expect if the children were at Lawrence Kohlberg's first stage of moral reasoning and that the apparently altruistic needs-oriented explanations have often – and presumably unfairly – been coded as stage 2, that is, as an immature, "preconventional" way of thinking about moral problems (see Nancy Eisenberg-Berg and Michael Hand, "The Relationship of Preschoolers' Reasoning About Prosocial Moral Conflicts to Prosocial Behavior," *Child Development*, vol. 50, 1979, pp. 356-63).

8. The tendency to define *altruism* so narrowly that only Mother Teresa would qualify for the label both reflects and perpetuates a cynical view of human nature. It would never occur to us to define *aggression* so as to exclude everything short of mass murder.

*"I typed my report, but I had to use a lot of Liquid Paper."*

9. Kohn, Ch. 3.

10. Norma Deitch Feshbach and Seymour Feshbach, "Affective Processes and Academic Achievement," *Child Development*, vol. 58, 1987, pp. 1335-47. For more research on cognitive skills and perspective-taking, see David W. Johnson and Frank P. Johnson, *Joining Together: Group Theory and Group Skills*, 3rd ed. (Englewood Cliffs, N.J.: Prentice-Hall, 1987), p. 244.

11. For example, see David Johnson et al., "Effects of Cooperative, Competitive, and Individualistic Goal Structures on Achievement: A Meta-Analysis," *Psychological Bulletin*, vol. 89, 1981, pp. 47-62; David W. Johnson and Roger T. Johnson, *Cooperation and Competition* (Edina, Minn.: Interaction Book Co., 1989), especially Ch. 3; and Robert E. Slavin, *Cooperative Learning: Theory, Research, and Practice* (Englewood Cliffs, N.J.: Prentice-Hall, 1990), especially Ch. 2.

12. Thomas Gordon, *Teaching Children Self-Discipline* (New York: Times Books, 1989), p. 7.

13. Robert J. Sternberg, "Prototypes of Competence and Incompetence," in Robert J. Sternberg and John Kolligian, Jr., eds., *Competence Considered* (New Haven: Yale University Press, 1990), p. 144.

14. Mark R. Lepper and David Greene, "When Two Rewards Are Worse Than One: Effects of Extrinsic Rewards on Intrinsic Motivation," *Phi Delta Kappan*, April 1975, pp. 565-66.

15. See, for example, Edward Deci and Richard Ryan, *Intrinsic Motivation and Self-Determination in Human Behavior* (New York: Plenum Press, 1985); Mark R. Lepper and David Greene, eds., *The Hidden Costs of Reward* (Hillsdale, N.J.: Erlbaum, 1978); and the work of John Nicholls, Teresa Amabile, Judith M. Harackiewicz, Mark Morgan, and Ruth Butler. I have reviewed some of this research in "Group Grade Grubbing Versus Cooperative Learning," *Educational Leadership*, February 1991, pp. 83-87.

16. Quoted in Mark Morgan, "Reward-Induced Decrements and Increments in Intrinsic Motivation," *Review of Educational Research*, vol. 54, 1984, p. 5.

17. Cathleen L. Smith et al., "Children's Causal Attributions Regarding Help Giving," *Child Development*, vol. 50, 1979, pp. 203-10; and C. Daniel Batson et al., "Buying Kindness: Effect of an Extrinsic Incentive for Helping on Perceived Altruism," *Personality and Social Psychology Bulletin*, vol. 4, 1978, pp. 86-91.

18. Miron Zuckerman, Michelle M. Lazzaro, and Diane Waldgeir, "Undermining Effects of the Foot-in-the-Door Technique with Extrinsic Rewards," *Journal of Applied Social Psychology*, vol. 9, 1979, pp. 292-96.

19. Richard A. Fabes et al., "Effects of Rewards on Children's Prosocial Motivation," *Developmental Psychology*, vol. 25, 1989, pp. 509-15.

20. Suzanne Daley, "Pendulum Is Swinging Back to the Teaching of Values in U.S. Schools," *New York Times*, 12 December 1990, p. B-14.

21. Quoted in David Hill, "Order in the Classroom," *Teacher Magazine*, April 1990, p. 77.

22. Joan E. Grusec and Theodore Dix, "The Socialization of Prosocial Behavior: Theory and Reality," in Carolyn Zahn-Waxler, E. Mark Cummings, and Ronald Iannotti, eds., *Altruism and Aggression: Biological and Social Origins* (Cambridge: Cambridge University Press, 1986), p. 220.

23. Joan E. Grusec et al., "Modeling, Direct Instruction, and Attributions: Effects on Altruism," *Developmental Psychology*, vol. 14, 1978, pp. 51-57.

24. Joan E. Grusec and Erica Redler, "Attribution, Reinforcement, and Altruism: A Developmental Analysis," *Developmental Psychology*, vol. 16, 1980, pp. 525-34.

25. Angelo Strenta and William DeJong, "The Effect of a Prosocial Label on Helping Behavior," *Social Psychology Quarterly*, vol. 44, 1981, pp. 142-47.

26. See James H. Bryan and Nancy H. Walbek, "Preaching and Practicing Generosity," *Child Development*, vol. 41, 1970, pp. 329-53; James H. Bryan and Perry London, "Altruistic Behavior by Children," *Psychological Bulletin*, vol. 72, 1970, pp. 200-211; Martin L. Hoffman, "Altruistic Behavior and the Parent-Child Relationship," *Journal of Personality and Social Psychology*, vol. 31, 1975, pp. 937-43; and Marian Radke-Yarrow, Phyllis M. Scott, and Carolyn Zahn-Waxler, "Learning Concern for Others," *Developmental Psychology*, vol. 8, 1973, pp. 240-60.

27. Ben Spiecker, "Psychopathy: The Incapacity to Have Moral Emotions," *Journal of Moral Education*, vol. 17, 1988, p. 103.

28. For an analysis of the harms of competition in the classroom and elsewhere, see Alfie Kohn, *No Contest: The Case Against Competition* (Boston: Houghton Mifflin, 1986).

29. David W. Johnson et al., "Effects of Cooperative Versus Individualized Instruction on Student Prosocial Behavior, Attitudes Toward Learning, and Achievement," *Journal of Educational Psychology*, vol. 68, 1976, pp. 446-52; and Bette Chambers, "Cooperative Learning in Kindergarten: Can It Enhance Students' Perspective-Taking Ability and Prosocial Behaviour?," unpublished manuscript, Concordia University, Montreal, 1990.

30. See, for example, the research cited in David W. Johnson and Roger T. Johnson, "The Socialization and Achievement Crisis: Are Cooperative Learning Experiences the Solution?," in Leonard Bickman, ed., *Applied Social Psychology Annual 4* (Beverly Hills, Calif.: Sage, 1983), p. 137; and Elliot Aronson and Diane Bridgeman, "Jigsaw Groups and the Desegregated Classroom: In Pursuit of Common Goals," *Personality and Social Psychology Bulletin*, vol. 5, 1979, p. 443.

31. Another classroom management issue is raised by Carolyn Zahn-Waxler. She warns that a teacher who routinely and efficiently takes care of a child in distress in order to preserve order in the classroom may unwittingly be teaching two lessons: 1) that "people do not react emotionally to upset in others" and 2) that, "if someone is hurt, someone else who is in charge will handle it" ("Conclusions: Lessons from the Past and a Look to the Future," in Zahn-Waxler, Cummings, and Iannotti, p. 310).

32. For more about the Child Development Project, see Alfie Kohn, "The ABC's of Caring," *Teacher Magazine*, January 1990, pp. 52-58; and idem, *The Brighter Side of Human Nature*, Ch. 6. For accounts written by members of the staff, see Victor Battistich et al., "The Child Development Project: A Comprehensive Program for the Development of Prosocial Character," in William M. Kurtines and Jacob L. Gewirtz, eds., *Moral Behavior and Development: Advances in Theory, Research, and Applications* (Hillsdale, N.J.: Erlbaum, 1989); and Daniel Solomon et al., "Cooperative Learning as Part of a Comprehensive Classroom Program Designed to Promote Prosocial Development," in Shlomo Sharan, ed., *Cooperative Learning: Theory and Research* (New York: Praeger, 1990).

# III-6

# Preparing for Idaho's Future

PAUL BERMAN
BW ASSOCIATES

DRIEK ZIRINSKY
BOISE STATE UNIVERSITY, IDAHO

Reprinted with permission from Paul Berman.

# PREPARING FOR IDAHO'S FUTURE

## An Action Plan
## for
## Education in the 21st Century

# SUMMARY

R-117/1

**Project Staff**
*Paul Berman Driek Zirinsky*

*Phyllis Edmundson*
*Barry Bull*
*Janice Fletcher*
*Gary Lyman*
*Wayne Phillips*
*Richard Sagness*
*Bob Smith*
*Carolyn Thorsen*
*Veronica Daley*
*Chris Gleason*
*Greg Martinez*
*Donelle Ruwe*
*Betsy Williams*

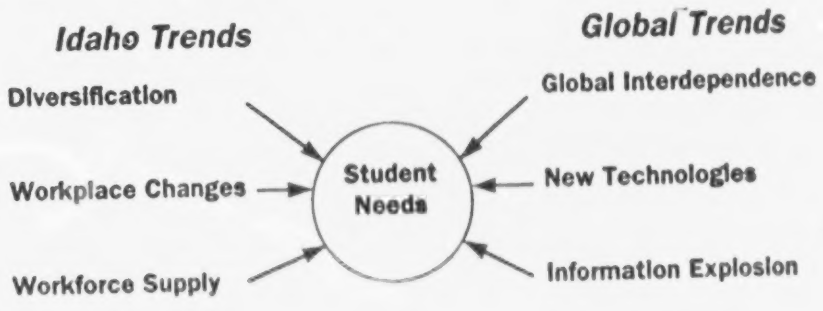
# FUTURE STUDENT NEEDS

To find out what Idahoans believe the needs of students will be in the 21st Century, we conducted an investigation throughout the state, one which involved students, parents and teachers, school administrators, and community and business leaders. The Project selected ten communities to represent the diversity of Idaho and of its school districts and visited each of these school communities. In addition, we systematically interviewed people responsible for personnel at 24 representative businesses and held panel sessions in three regions of the state about Idaho's future employment trends. Among all these Idahoans — numbering close to a thousand people — a broad consensus emerged about the future needs of Idaho's students.

### Challenges of the Future

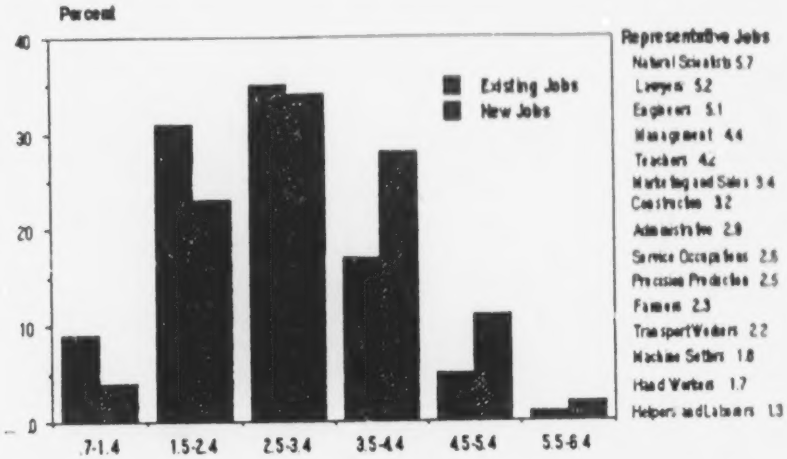
World, regional and local forces affecting Idaho's future are great and, to some extent, unpredictable. Yet major trends can be discerned and prepared for. Our meetings with employers and citizens throughout the state confirmed the all-too-frequent newspaper headlines: the larger global economic movement toward interdependence, the advance of new technologies, and the information explosion are affecting Idahoans whether in large or small businesses, in rural or more urban settings. The message for the future is clear: in Idaho, as across the nation, what young people will have to learn in school is changing. In their future roles as citizens and as workers, they will need higher levels of competence to function well.

### Forces Affecting Student Needs



One reason the expectations for young people have changed is that the workplace itself has changed dramatically in the last 25 years. To prepare young people for work, schools must anticipate the new workplace. Many jobs today, even those that do not require post-secondary education, do require a high level of competence, knowledge and skill perhaps matching the level of expectation held in the past only for those going to college.

### *Low Skilled Jobs Are Declining Nationally*



Previously, employers wanted young people to be obedient order-followers. Today, workers are expected to think for themselves, employ sophisticated problem-solving skills, and know how to work in collaboration with others. The structures of most classrooms, however, still reinforce the older values and notions, not the newer ones.

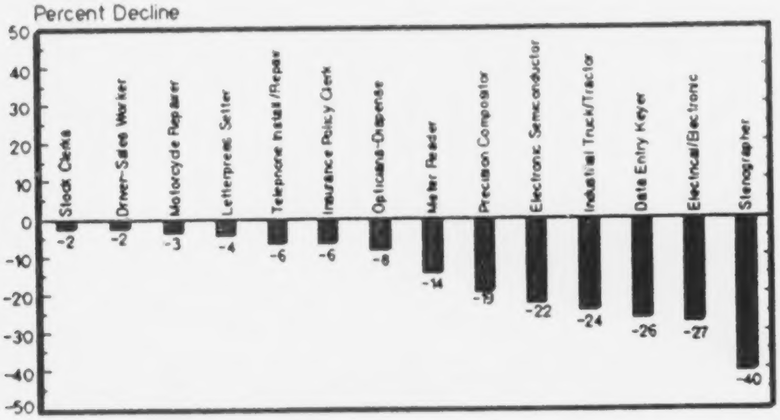
A young person can anticipate a new role as a citizen in our democracy, a role that also requires escalating levels of knowledge and ability. Public policy issues increasingly require voters who can understand scientific concepts. Some examples include environmental and waste disposal issues; water and forest management; and the rate of development of Idaho's wild lands. Being an informed voter today necessitates a higher reading level than in the past.

### *The "New Basics"*

Virtually all people we interviewed — from executives to entry-level employees, from farmers to computer experts, from students to teachers — thought that schools should be expected to provide youth with a sound education leading to skills and abilities people commonly refer to as "the basics." But this expectation has a new twist for the future: basic skills will no longer be rudimentary readin', writin' and 'rithmetic; tomorrow's graduates will be expected to be literate across a variety of areas relevant to a productive life in an increasingly competitive world.

## What Idaho Employers Say

In particular, Idahoans expect that young people will be literate with language, able to read, write and communicate clearly and fluently in a variety of settings. Increasingly, the global economy is becoming a daily practical reality for many Idahoans, as citizens and as workers. Each will need to develop understanding and awareness of languages, cultures and outlooks fundamentally different from those he or she grew up with. Idahoans we interviewed were unanimous in considering global knowledge and awareness as a basic skill for all students in the future, and many felt that the use of languages other than English would be essential.



**Occupations on the Decline in Idaho**

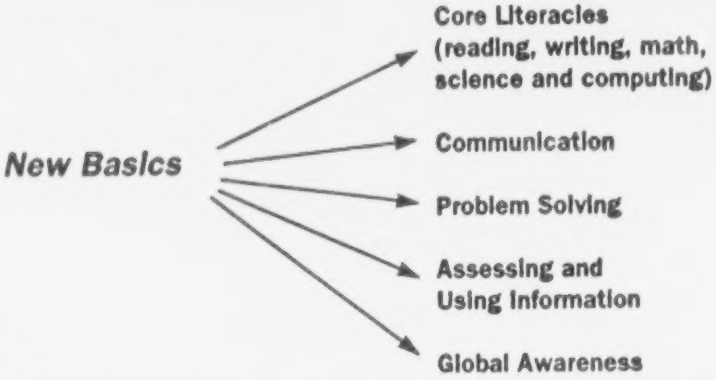
Idahoans also expect that young people will be literate in science and math. In this context, literacy means having a level of knowledge of scientific and mathematical concepts which goes beyond the old notion of basic math — the simple one-step calculations, math for consumers, or workplace math. Students will need to be comfortable with mathematical and scientific reasoning and be able to use them for problem-solving in both work and life situations.

According to every Idahoan we interviewed, students will also need to be computer literate. They should have keyboard skills and basic familiarity with the computer as a tool for doing tasks. They should have working familiarity with the computer, robotics, and telecommunications technologies that are rapidly becoming staples of contemporary life.

BEST COPY AVAILABLE

At an even more fundamental level, Idahoans expect that youth will have the ability to solve problems and know how to reason through complex steps. Some of our respondents described this as "critical thinking ability."

***Work will become more participative and employees will be more diverse in gender, age, and ethnicity.***

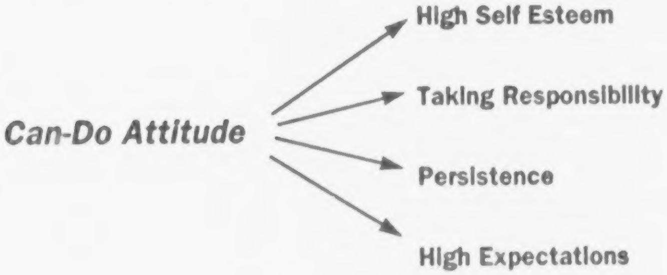

New Basics
Core Literacies (reading, writing, math, science and computing)
Communication
Problem Solving
Assessing and Using Information
Global Awareness

It was also widely noted that students will need the ability to access information, not to be crammed full of facts. In this age of information explosion, students need strategies for finding the information they need when they need it.

***Can-do Attitude***

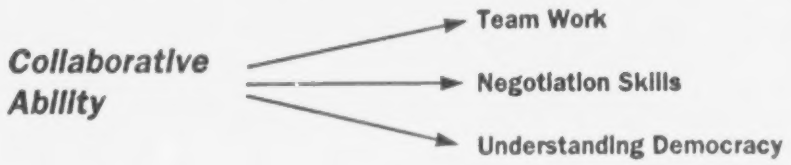
Employers and workers alike agreed that, to get along in the world, school finishers will need a set of positive characteristics. They will need to hold high expectations of themselves which match those held by their families, by the schools, and by the society. A strong sense of self-esteem, confidence and competence were mentioned frequently: in short, a "can-do" attitude. Many employers and community leaders said students also will need the ability to deal with set-backs, discouragements, and failures.

***Increasing competition places a premium on productive employees having a positive attitude.***


Can-Do Attitude
High Self Esteem
Taking Responsibility
Persistence
High Expectations

### *Collaborative Ability*

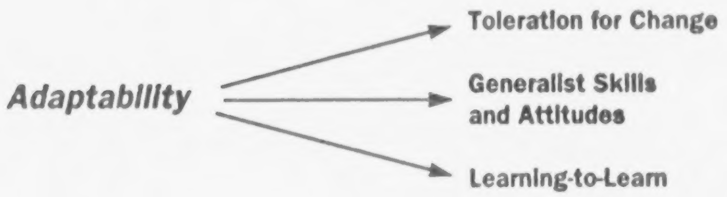
Each student will need to leave school with collaborative ability — the ability to be a team worker, a person who can cooperate on a task. Employers gave us many concrete examples of how fewer and fewer people work in isolation in today's workplace — and they expect the trend toward close collaboration to become more pronounced. In personal lives, the increasing stress of multiple roles that people fill in Idaho community life will require more finely honed skills in negotiation and deeper understanding of the principles of democracy.


Collaborative Ability
Team Work
Negotiation Skills
Understanding Democracy

***Today's entry-level employees are often deficient in basic skills, and many professional jobs are filled from outside Idaho. Tomorrow's workers need more comprehensive and higher order skills.***

### *Adaptability*

Finally, educators and older citizens of Idaho reminisced about how life was simpler not too many years ago and how quickly things were changing. They said that, in the face of this rapid change, young people will need the ability to adapt as a routine part of successful living. Adults of the future will need a high level of tolerance for the increasing diversity that is slowly coming to Idaho. They will need to have learned strategies for coping with change, as well as the ability to be lifelong learners.


Adaptability
Toleration for Change
Generalist Skills and Attitudes
Learning-to-Learn

***Besides a shift to service sectors, jobs and workplaces will change rapidly. Citizenship will require active involvement and diverse experience.***

### III-7

# The New Definition of Learning: The First Step for School Reform

BEAU FLY JONES AND TODD FENNIMORE
NORTH CENTRAL REGIONAL EDUCATIONAL LABORATORY

From *Restructuring to Promote Learning in America's Schools: A Guidebook* (Oak Brook, Ill.: North Central Regional Educational Laboratory), pp. 10–18. Reprinted with permission of the North Central Regional Educational Laboratory.

# What Is the New Definition of Learning?

## "Old" and "New" Definitions of Learning

Educators know when they see meaningful learning in a classroom. The students are actively engaged in challenging tasks that have applications to their lives. They do not settle for rote understandings. They truly understand the subject matter because they can apply their knowledge to new situations. They see learning as an integral and enriching part of their everyday lives, something highly enjoyable and intrinsically motivating. Their teacher or teachers flow in and out of their activities, sometimes listening, sometimes redirecting their thoughts, sometimes coaching or explaining.

This vision of learning is not "new" in two respects. First, successful learners of today probably have many of the same attributes as those of yesterday. Second, recent research on learning reflects common sense visions of learning with historical roots in earlier philosophies of education such as those of Socrates, Dewey, Piaget, and Vygotsky. These earlier philosophers/ educators shared a rich vision of learning. Love of learning was nourished by teachers. The locus of learning was the learner, and the potential for what children could learn was a frontier.

## The New Definition of Learning: Overview

Today, the dominant metaphor for learning in most schools is industrial. This suggests that the business of learning is additive and largely controlled by persons outside the learner. Schooling becomes an assembly line process where the objectives define what the end product of learning will be. The learner is a recipient. The teacher is an information giver. The principal is a manager who oversees the building, equipment, and people involved in the production process.

In contrast, recent research has generated some new terms such as meaningful learning, problem situated learning, reflective education, and cognitive apprenticeship. These terms suggest again that the essence of successful learning is internal to the individual, group, or community. The learner works hard to link the new information to prior knowledge and so constructs mental models and explanations. Further, the context of learning has come to be described as collaborative, diversified, and rich in resources and opportunities for learning.

Successful learning necessarily involves authentic tasks, objectives that engage the mind, and a focus on using the information that is acquired.

These new descriptions of learning are important politically. In the past, it was generally assumed that only the brightest of our nation's youth would be likely to have the attributes of successful learners, and that precious few public schools could provide an optimal instructional environment to foster successful learning for the majority of youth. Now, more and more people assume that research can help us identify the characteristics of successful learning and learning environments, that the performance of low-achieving students can be significantly modified, and that schools can make a positive difference. Equally important, schools can use current research to work toward providing better learning environments for adult members of the learning community. This includes providing support services needed for students and their families for meaningful learning to take place.

Current research is providing a broader view of learning. A number of research centers and researchers are studying how children learn and what helps and hinders their learning. In this video conference, three bodies of research are featured that help contribute to a broader definition of learning: philosophical, cognitive, and multicultural. Together, they help define learners philosophically, intellectually, and culturally. That is, meaningful learning takes place when teachers and students are aware of all three of these dimensions of learning. Through visits to actual school settings, viewers are able to see how learning is nourished in these three areas in various real-world settings.

## Research on Meaningful Learning

### Philosophical Perspectives

Philosophical perspectives focus on how learners develop assumptions about what sorts of persons they want to become and what sort of world they want to inhabit. Fundamental to philosophical thinking is knowing how to verify the validity of available information and develop criteria for making judgments that are essential to thinking. Equally important is developing the disposition to challenge and defend foundations of beliefs and judgments.

## A Focus on Higher Order Thinking

Philosophical reflection engages many higher order thinking skills or strategies. Indeed, much effort in the thinking skills movement focuses on defining principles of logic and fallacies of reasoning such as making generalizations without adequate information or with biased information. Accordingly, philosophers tend to develop extensive taxonomies of strategies and fallacies. It is this aspect of philosophical thinking that many educators refer to as critical thinking. Yet this is only one aspect of philosophy. Philosophers increasingly stress the importance of creative thinking such as defining a problem in a new way or changing the problem, looking at an issue from an unusual perspective, and internalizing the process of reflective inquiry to apply to all subject areas.

## Community of Inquiry

A philosophical perspective on learning also stresses the importance of creating a community of inquiry in the classroom. Students in such communities expect each other to be thoughtful and reflective, listen to each other with respect, reflect and build on one another's ideas, demand evidence to support opinions, assist each other in drawing out the implications of what has been said, identify and challenge the assumptions of arguments, and question one another's definitions. Indeed, some philosophers argue that such inquiry is a moral imperative in a democratic society. Without it, we are prisoners of our own beliefs, unable to judge right and wrong with integrity, unable to make informed judgments. Philosophers want learners to be philosophical in the strong sense with a passion for reflection and inquiry pervading their thoughts and their lives.

## Examples

In considering examples of philosophical thinking, it is vital to understand that engaging students in classifying activities or having them identify fallacies of reasoning in prefabricated materials is not philosophical thinking in the strong sense. Rather, examples of philosophical thinking involve making observations, classifying, challenging premises, and the like for the purpose of making judgments about oneself or the world. Thus, students might compare and contrast people and robots for the purposes of understanding something about themselves as human beings. What is vital is that the students puzzle and wonder about the essence of humanity.

Presenter:
Matthew Lipman

While most people associate such intellectual moves with college level philosophy classes, Matthew Lipman, Director, Institute for the Advancement of Philosophy for Children, has helped elementary children learn how to engage in philosophical discourse with their peers. Lipman's extensive research has established that even very young students can attain philosophical understanding that the world is ambiguous and mysterious, and that purposeful dialogue with others can help illuminate the "blooming, buzzing universe." Further, NCREL believes that Lipman's vision of learning as a community of inquiry is important for schools to consider as they fashion their own visions of learning. (See Presenters' Biographical Information on page 48.)

## Cognitive Perspectives

Cognitive perspectives on learning come from cognitive psychology, educational psychology, psycholinguistics, computer sciences, expert teaching, and related fields. Although these perspectives developed largely independently of philosophical perspectives, both include a focus on higher order thinking, including strategic thinking or metacognition, collaborative learning, and a passion for inquiry. They differ primarily in focus of inquiry and the methods of study associated with respective lines of inquiry. Whereas philosophers strive to define what to do or believe, cognitive perspectives focus on such questions as understanding how students acquire and use knowledge in disciplinary contexts such as math and reading, what is involved in self-regulated learning, and what conditions produce optimal learning.

### Importance of Knowledge

Cognitive perspectives of learning emphasize that proficient learners have extensive subject matter knowledge. Proficient learners use this content knowledge to interpret new situations, define and solve problems, make judgments, classify information, and reason. Further, they represent and organize what they know in terms of patterns and principles. They are aware when the disciplinary knowledge conflicts with their prior knowledge, and they strive for conceptual change. These and other behaviors make their knowledge assessable; they can apply their knowledge to new situations. Others emphasize that proficient learners develop a repertoire of thinking/learning strategies, including strategies for critical and creative thinking and metacognitive strategies for monitoring their learning.

## Self-Regulated Learning and Strategic Thinking

Self-regulated learning from the cognitive perspective involves not only setting learning goals, but also monitoring the process of learning, thinking strategically about how to accomplish a task, assessing what is learned, and determining the next steps for learning. Even young children are capable of self-regulated learning. Children in the Perry Preschool Project, for example, begin their days making decisions about what they would like to accomplish during the day, monitor how well they implement plans, make appropriate modifications, and assess their learning at the end of the day. Proficient learners also develop a repertoire of learning strategies such as elaborating and summarizing, as well as strategies for critical and creative thinking.

## Authentic Tasks and Context

Cognitively demanding learning requires more than recalling facts, reciting definitions, or performing rote calculations. According to cognitive perspectives, the essence of learning is linking new information to prior knowledge and making connections to real-world situations or contexts. In fact, cognitive psychologists increasingly emphasize this contextual aspect of successful learning, arguing that learning must be "situated" in authentic tasks such as having real audiences for one's writing and applying a concept or process to solve a real-world problem. Where possible, tasks in school should be aligned with tasks and experiences that students encounter out of school. In this regard, dialogue, including self-questioning and collaboration with others, is critical in the process of meaningful learning.

## Examples

A classroom with a thought-provoking atmosphere is "abuzz" with students sharing ideas, solving problems, probing an issue, exploring an argument, or looking at different perspectives. Children at Sabin Elementary School in Chicago, for example, have the opportunity to describe their experiences at the Science Museum to specially trained mothers serving as scribes. Later these scribes will work with the children to convert the stories into brightly colored books to share with others in the class. Or consider a high school example. At Harbison High School in Orlando, Florida, students explore the meaning of literature through a historical perspective. Before studying each work of literature, students learn the history of the period as they learn the characteristics of the genre. When they read the text, they relate the content back to the historical setting.

### Learning Environments for Low-Achieving Children

Cognitive psychologists generally believe that attributes of successful learning such as strategic thinking and self-regulated learning can be taught to low-achieving students, given appropriate learning environments. There is considerable debate about exactly what conditions maximize learning for low-achieving students. However, most agree that providing such students with authentic tasks and problems with real-world applications, high expectations, and interaction with teachers and peers who are themselves good thinkers may have substantial benefits. They also emphasize the importance of dialogue with others. It is through dialogue that parents, teachers, and peers model and mediate learning. It is also through dialogue and self-dialogue that students formulate hypotheses and questions, puzzle, establish criteria, make a wide range of choices and judgments, negotiate settlements, and resolve disputes.

### Presenter:
### Lauren Resnick

Lauren Resnick, Director, Learning Research and Development Center, University of Pittsburgh, has spearheaded much of the research in the cognitive perspective. Among other things, her vision of learning seeks to align activities in schools with real-world tasks and learning environments outside schools. Toward this end, she emphasizes the concept of cognitive apprenticeships developed by Allan Collins, John Seeley Brown, and others. According to this concept, meaningful learning occurs when students tackle the complex tasks engaged in by adults in a given area, gradually building expertise with experience; meaningful learning also occurs when teachers serve as coaches, at first highly supportive, then gradually encouraging independent learning. (See Presenters' Biographical Information.)

## Multicultural Perspectives

Multicultural perspectives both reflect and extend philosophical and cognitive perspectives. Recent research focusing on multicultural education is based on two fundamental assertions: (1) all students can learn, and (2) students learn best when their learning tasks are built on real experiences and meanings from the students'own lives. Learning is an active process of inquiry, and the construction of meaning must begin with what the child already knows. What is distinctive about multicultural perspectives is the urgency of applying these principles to non-dominant or minority children. Too often schools do not legitimize the knowledge or experiences these children bring to school. Instead, schools are most likely to label these childen as failures because their backgrounds — usually their language and culture — are seen as inadequate preparation for learning. According to multicultural perspectives, every culture brings habits of thought, resources, and contexts which have built into them vehicles that promote learning and inquiry. Accordingly, children of any culture can and should have curriculum and instructional practices that draw from that culture.

### The Need to Build on Background Knowledge

Multicultural perspectives insist that language and culture play critical roles in learning. In the past, old models of learning tried to erase or ignore a minority child's cultural and linguistic background. Today's researchers, however, realize that this background represents the cornerstone of the child's prior knowledge. It is the foundation upon which all other learning is built. It is therefore crucial to link new learnings to the culture and experiences of the students. This means not only respecting and building on the knowledge that children bring to school as legitimate contributions to classroom dialogue, but, also working hard to use the familiar experiences of students as the starting point for learning concepts and principles.

### Implications for School Reform

This line of research has some important implications for school reform. First, schools are presently organized so that one's cultural background largely determines one's access (or lack of access) to educational opportunities and resources. For many non-dominant or minority children, this organization means tracking in low-level classes and virtual exclusion from high-level content and instruction. If schools restructure so that their curricula and instruction build on the language and culture of all children, these exclusionary and discriminatory practices will end. This means that schools must pay more than lip service to the inclusion of children of all groups in thoughtful learning activities.

Second, schools must make multicultural education a part of the learning objectives for all students. Belief in the efficacy and value of diverse cultures — not just those of the dominant group — must pervade all curricula and all instructional practices. This means making both students and educators aware of the strengths of diverse cultures and explicity teaching students to value individual differences in the classroom. However, changes such as these require a complete re-thinking of the school culture and organizational structure, as well as of learning objectives and instructional techniques.

Presenter:
Hubert Dyasi

Hubert Dyasi, Director of the City College Workshop Center in New York, is one science researcher who has linked multicultural research to the reform of teaching and learning in schools with a focus on science. Following the work of his colleague Lillian Weber, his program at the Workshop focuses on restructuring classroom time and space so that children are actively engaged with their local environment in Harlem, and instruction builds on what the children already know.

Common materials and technologies (elevators, bicycles, buildings) and ordinary observable phenomena (swinging, the smell of sea water, blowing of soap bubbles) become entry points for children to learn science. Additionally, Dyasi argues that concepts and principles are accessed through language use and that minority children must be proficient in "the language of currency." Thus, a major task of school reform is to promote dialogue about the phenomena under study. (See Presenters' Biographical Information.)

## Some Research-Based Assumptions About Learning

The three bodies of research on learning featured in this first video conference help broaden current definitions of learning. Often what students are thought to learn or master in traditional definitions of learning is subject matter. Yet we know that all students, regardless of age, come to learning experiences with cognitive, philosophical, and cultural strengths and needs. They also learn about themselves through reflection, about how to think about others different from themselves, and about learning itself.

Learners and teachers who are aware of these three dimensions of learning come to the educational process with a different set of assumptions, compared to those with traditional views of learning. They understand that successful learners are different from less proficient learners in important ways that are learned through social and cultural experiences inside and outside the classroom.

### Characteristics of the Successful Learner

When we look back at what each perspective has to say about successful learning, four characteristics are outstanding: successful learners seem to be knowledgable, self-determined, strategic, and empathetic. These characteristics form a **beginning list** of research-based assumptions about meaningful learning.

**Knowledgable** Successful learners have amassed substantial knowledge of content and can use it to think with fluency and authority. They can define and solve problems critically and creatively. They have a strong sense of what they believe and why, and they constantly evaluate the quality of information they receive and produce. They devote substantial time to reflecting and puzzling. They are aware of the strengths and needs of others, including those different from themselves. They value diversity and work hard to be fair-minded.

**Self-Determined** Successful learners consciously use and expand the tools they have to engage in meaningful learning. They are highly motivated and feel they have the power to promote their own development. They feel good about themselves as learners, confident that they can succeed. They have firm beliefs about the value of hard work and the effects of it on their success. Successful learners persevere in the face of difficulties. Successful learners also regard the world as full of opportunities to learn. They consider choices, examine reasons, and observe the consequences of their actions.

**Strategic** Successful students have a repertoire of effective strategies for learning various subjects, for thinking about and controlling their own learning, for detecting errors and fallacies in their thinking and the thinking of others, for problem solving and decision making, and for thinking creatively. Using these strategies, they can organize what they know and construct mental models, plan their study time, and decrease anxiety. They can monitor their comprehension and problem solving, constantly pausing to compare the new information to prior knowledge, check what they have learned and what they need to review, and summarize. Successsful students learn to orchestrate these strategies in a dynamic flow as they move in and out of different tasks and phases of learning.

**Empathetic** Successful students often recognize that much of their success involves their ability to communicate with others. However, they are also able to view themselves and the world through the eyes of others. This means not only examining one's beliefs critically, including beliefs and judgments about the self, but also examining the beliefs and circumstances of others, keep ing in mind the goal of enhanced understanding and appreciation. These interpersonal skills are particularly important in understanding and appreciating other cultures. Meaningful learning involves learning how to identify the strengths of others as well as to be supportive in give-and-take relationships. Successful students value sharing experiences with persons of different backgrounds as enriching their lives.

# III-8

# "Expect the Best: From Our Students, of Our Staff, for Our Community"

SNOQUALMIE VALLEY SCHOOL DISTRICT NO. 410, SNOQUALMIE, WASHINGTON

## Mission Statement

It is the mission of the Snoqualmie Valley School District to meet the individual learning needs of its students, thereby enabling them: to identify and realize their potentials, to develop skills and attitudes compatible with life-long learning, and to be knowledgeable and productive citizens.

To accomplish our mission, the Snoqualmie Valley School District will:

**Emphasize** basic skills, effective communication, information processing, responsible decision making, problem solving, critical and creative thinking;

**Encourage** the highest and greatest development of intellectual and creative abilities through a diverse curriculum, varied teaching techniques and individualized learning in a challenging, stimulating and caring environment.

**Foster** acceptance, appreciation and respect among students for themselves and others in the tradition of the cultural and ethnic diversity of our community and nation;

**Promote** understanding and appreciation for the rights, duties, privileges and responsibilities of citizenship;

**Prepare** students to live and work in a diverse society and everchanging world;

**Encourage** the parent-student-staff learning partnership;

**Involve** parents, students and staff in the decision-making process in matters affecting the quality of education provided in our schools.

---

Reprinted with permission from the Snoqualmie Valley School District #410, Snoqualmie, Washington.

# III-9

# "Learning for a Lifetime": Mission and Student Goals

SALEM-KEIZER PUBLIC SCHOOLS, SALEM, OREGON

## Our Mission

In partnership with the community, we will insure that each student has the essential knowledge, skills and attitudes to be a lifelong learner, a successful family member, a contributing citizen, and a productive worker in a changing and increasingly diverse world.

## Student Goals

We will focus all efforts to help each student develop the necessary knowledge and experiences to:

- read and communicate effectively
- think critically and solve problems
- develop positive self concept, respect for others, and healthy patterns of behavior
- work effectively in groups as well as individually
- develop creativity and show appreciation for the arts
- demonstrate civic, global, and environmental responsibility
- recognize and celebrate diversity
- exhibit technological literacy

## Achieved Through . . .

### *Active Partnerships for Learning*

Our partnerships with students, families, businesses, and the broader community will:

- provide for the safety and welfare of all children
- be driven by a commitment to public education
- focus on student needs
- make full use of community resources for learning
- operate from a spirit of mutual respect, trust, and interdependence
- be clear about roles and foster meaningful involvement
- continue to increase in variety and effectiveness

### *Organizational Effectiveness*

As an organization, the School District will:

- focus on student outcomes and total quality of all services
- be a learning organization that continuously assesses and improves
- have high levels of teamwork and trust
- involve people in making decisions that affect them
- treat people fairly and with equity
- recognize and value diversity among people and cultures

---

Reprinted with permission from Salem-Keizer School District, Salem, Oregon.

- refuse to tolerate racism, discrimination, harassment and/or prejudice
- recognize and reward growth
- continue to develop and support a high quality staff
- actively seek and value input from throughout the community
- provide a welcoming, open door environment for parents and community

# III-10

## Shared Meaning: Student Learning Goals

Aurora Public Schools, Aurora, Colorado

**EMPHASIZE** basic skills, effective communication, ability to access and process information, responsible decision making, problem solving, critical thinking, and creative endeavors.

- Early and continued emphasis on reaching a high level of literacy including decoding skills, comprehension and fluency in order to read for information and pleasure
- To accomplish new tasks using effective verbal and non-verbal communication, with various audiences.
- Effectively communicate observations, impressions and ideas.
- Develop the habit of analyzing, sorting, evaluating written, verbal and electronic information as a continuous learning process.
- Interlace, transfer and apply basic skills, concepts, problem solving strategies and decision making skills throughout the disciplines of learning.
- Demonstrate a broad understanding of local, state, national, and world systems and cultures.
- Develop self-responsibility in applying the components of healthy living.
- Exposure to and opportunities for participation in all areas of the arts.
- Enhance memorization strategies to support higher order thinking skills.
- Analyze the long term consequences of short term actions.
- Generate and respect creativity.
- Read for information: utilizing specific instructions to accomplish new tasks.
- Read for response: communications, impressions, criticisms, and feelings.
- The ability to recognize the intention of a speaker and to be aware of the techniques a speaker is using.
- Develop and evaluate the logic of an opposing point of view.
- Interlacing of reading, writing, speaking, and listening in all disciplines.
- Provide opportunities to develop, apply and evaluate ideas with others in everyday living.
- Students will use technology to enhance personal production, transmit, access and retrieve information.

**EXPECT** the highest and greatest development of self-esteem, intellectual, and creative abilities through a diverse and integrated curriculum using varied teaching techniques and resources in a challenging, stimulating, caring, and safe environment.

- Develop a positive sense of self worth.

---

Reprinted with permission from the Aurora Public Schools, Aurora, Colorado.

- Take responsibility for and manage own learning time and effort.
- Ability to set short and long term goals.
- Evaluate own progress toward achieving goals.
- Demonstrate sensitivity to the views of others.
- Take pride in work and demonstrate a positive work ethic.
- Demonstrate high expectations for own productivity.
- Become confident in one's own academic ability.
- Develop a sense of personal effectiveness.
- A sense of self esteem which engenders success in work.
- Recognize one's own strengths.
- Build and maintain the esteem of self and others.
- Develop a belief in one's ability to shape the future.
- Display honesty, integrity, and courage.
- Is flexible and patient.
- Work effectively as a member of a team.

**FOSTER** acceptance, appreciation, and respect among students for themselves and others in the tradition of the cultural and ethnic diversity or our community, nation, and world.

- Recognize that each person is unique and valuable in their diversity.
- Understand how actions and attitudes affect others.
- Recognize common stereotypes and how they affect thinking.
- Acquire a wide exposure of different cultures and their contributions through a total multi-cultural curriculum which includes languages.
- View and interpret facts, events, concepts, and theories from varying perspectives.
- Understand, investigate, and determine how implicit cultural assumptions, frames of references, perspectives, and biases within a discipline influence how knowledge is created.
- Acquire skills to deal with differences and conflict.

**PROMOTE** understanding, and appreciation of the rights, duties, privileges and responsibilities through active involvement as a citizen in an everchanging world.

- Illustrate an understanding of day to day rights, duties, and responsibilities that exist as a citizen in economic, social, and political systems.
- Demonstrate leadership through consensus building, organizational skills and vision.
- Willingness to participate and make a difference in the world we all share.
- Demonstrate social responsibility through volunteering in some form of community service.

**PREPARE** students to live and work in a rapidly changing technological future.

- Understand and select appropriate technology and use its applications.
- Apply technology to solve relevant life problems.
- Acquire work place literacy.
- Use technology ethically and humanely.

**CULTIVATE** students' environmental awareness and responsibilities.

- Students will demonstrate environmental awareness and responsibility based on a fundamental grasp of basic subject areas.
- Environmental concepts will be included in basic subject areas such as math, reading, communication, social sciences, physical and natural science.
- Students will be able to evaluate consequences and make thoughtful choices through use of critical thinking, problem solving and effective decision making skills. Decision making on environmental issues will be based on science skills and an inquisitive open-mindedness.

# III-11

# Maine's Common Core of Learning: An Investment in Maine's Future

COMMISSION ON MAINE'S COMMON CORE OF LEARNING

Reprinted with permission of the Maine Department of Education, Augusta. (ED 327 509)

## THE COMMISSION'S JOURNEY

Since February 1989, the members of the Commission on Maine's Common Core of Learning have been on an educational journey. We began as a group of forty-five individuals from all walks of life and have evolved into a community of learners. We read widely, listened to and read about the current issues in each content area, and processed letters and public comment from eight regional forums. We listened to our thirty-eight student readers from three high schools. Debate and discussion of issues led us gradually to agreement about the knowledge, skills, and attitudes Maine students need by the time they graduate from high school. In defining this common core, we found it arranged itself in a way that will challenge Maine residents to reconsider how curriculum is organized and how schools are structured. Consequently, our report has the potential to revolutionize the way Maine students are educated.

Our early meetings focused on what skills and attitudes graduating high school students should possess to be productive citizens, what will be essential for them to know considering the demands the twenty-first century will place on them as adults. We heard reports on the current profile of Maine high school students and from the Maine Aspirations Compact, Human Resources Development Council, and the Commission on Maine's Future.

In later meetings, we listened as leaders in each of the subject areas presented current thought and research about those areas. This helped us add specificity to our base. As we gained insight into each subject, we began to see commonalities among them — knowledge, skills, and attitudes that overlapped the areas. When we viewed the concepts in this way, we found a dimension not seen when any one subject area was viewed in isolation.

We want our students to have this same view, to acquire the fundamental and specific aspects of each subject area and be able to integrate this knowledge to see the world as a multi-faceted whole.

# WHY WILL IT BE NECESSARY TO EDUCATE MAINE STUDENTS DIFFERENTLY FOR THE TWENTY-FIRST CENTURY?

*. . . the democratic promise of equal educational opportunity, half fulfilled, is worse than a promise broken. It is an ideal betrayed. — MORTIMER J. ADLER*[1]

Within the last forty-eight hours, in the towns, villages, and cities throughout Maine, approximately one hundred new Maine citizens have been born. They have entered a state rich in natural beauty and resources, a place remote in geography, yet inextricably linked to other parts of the nation and the world by technology.

We welcome these youngsters with a willingness to share the way of life we have known in Maine, yet with the realization that the economic and technological terrain of our state is changing rapidly. These children will enter the work force of the twenty-first century. Most of them will grow up to assume jobs that do not yet exist,[2] jobs that will be reshaped or invented to accommodate fast-paced technological development. They need the knowledge, skills, and attitudes that will prepare them to provide for themselves and others and will give them a future of options — personal and work related — in Maine and in the world at large. We, the residents of Maine, offer these children the promise of an education to help them achieve fulfilling work and to act responsibly in an increasingly complex society.

Current evidence suggests that some of our one hundred children will have little difficulty realizing our educational promise. Others, however, will be at risk — unable to give what is demanded by the conditions in our rapidly-changing world. Our promise to these children will be broken if schools do not prepare them for a lifetime of choices.

**We . . . offer these children the promise of an education to help them achieve fulfilling work and to act responsibly in an increasingly complex society.**

The challenges awaiting these children are not just economic. Their lives will be affected by global concerns of interdependent societies and cultures. These children will need to develop judgment, personal responsibility, and a sense of ethics as world citizens.

Maine, along with other areas of the United States and Canada, is faced with such intensifying environmental concerns as solid waste disposal and the effects of acid rain upon our timberlands. Many

nations have serious air, water, and land pollution. Our children, as part of the world community, will need to contend with these issues as well as the growing impact of the ozone depletion, global warming, and the destruction of tropical rain forests.

Our biotechnical wizardry, which will have saved several of our hundred children from starvation and such deadly diseases as smallpox, has created a host of moral and ethical problems for them. With a greater ability to control life itself, these children, as adults, will face such increasingly urgent questions as the right to die, the custody of frozen embryos, and the soaring national cost of health care. They will need to respond to the threat of nuclear weapons and nuclear waste, as well as to critical challenges like the AIDS pandemic, issues that will require scientific and political expertise from some and commitment and spiritual depth from all.

**They will need to respond to ... issues that will require scientific and political expertise from some and commitment and spiritual depth from all.**

Changing social factors at home will also affect the development of our hundred new Mainers. If current statistics hold, sixty of them will live in a home with only one parent by the time they are eighteen, many will spend large portions of time in child-care facilities, and at least twenty-five will know poverty in the important years between birth and age six.[3] Once in school, several of these children will come home and wait for adults in an empty house or apartment. At current rates, at least ten will still be functionally illiterate by the end of their teen years;[4] twenty-three will not complete high school;[5] and many more will manage to earn a diploma by drifting through a "General" course of study that will give them little preparation either for work or for further education. Twenty-five of our one hundred children will be chemically dependent or at risk of dependency by the time they reach adulthood.[6] Finally, one in fourteen of the girls will become pregnant as a teenager and several of our youngsters will become teen parents.[7]

As these typical one hundred children grow, they will enter an educational system that is struggling, often with commendable energy and dedication, to cope with watershed changes in our society. A century and a half ago, schools faced a similar challenge as a primarily agrarian society made the transition to the industrial age that has sculpted the economic might of the United States. Those schools produced a relatively small proportion of decisionmakers and a large mass of workers who were compliant, basically literate in English, and able to perform repetitive tasks effectively and in isolation — factory-model schools for an industrial economy. Until the 1950s, fewer than fifty percent of American students graduated from high school: our industries could absorb massive numbers of unskilled laborers.[8] Without much more than an eighth grade education, large numbers of Maine workers succeeded in valued traditional occupations — farming, logging, and working on the sea or at the mill.

Today, however, demographic trends, technological change, and tighter international competition have changed the employment landscape. Many opportunities in the mills and factories, once the mainstays of local economies, are disappearing as firms close plants, consolidate, or automate to survive. Since 1947, manufacturing jobs in Maine have declined from forty-four percent of the state's economy to twenty-two percent, and the decline continues as the service sector booms.[9] The smaller number of children born in Maine, and indeed in the nation, during the last two decades will also influence the futures of our new arrivals. Ninety percent of Maine's labor force of the year 2000 is already working.[10] Each year fewer young workers will enter the job market, and their required level of preparation for employment will rise as service and

**As these typical one hundred children grow, they will enter an educational system that is struggling, often with commendable energy and dedication, to cope with watershed changes in our society.**

information industries create virtually all the new jobs and most of the new wealth in our state. Maine employers will seek qualified workers not only among the young, but increasingly among women not now employed, the elderly, and people with special needs.[11]

Maine businesses, in need of literate workers with good problem-solving skills, are becoming alarmed at the lack of competency in reading, writing, and mathematics of entry-level workers.[12] Here and across the nation, industrial jobs are changing shape, depending more on computers and other sophisticated machinery, requiring workers with greater flexibility[13] and improved information management skills. With this increased demand for highly-skilled workers, the lack of such skills among the unemployed disqualifies most from these jobs.[14]

The emerging jobs in service and information industries need workers who have learned how to learn.[15] They need to know how to find and use information from libraries and other learning resources, not only for their employers' needs but for their own. Because

**Maine businesses ... are becoming alarmed at the lack of competency in reading, writing, and mathematics of entry-level workers.**

service work is also people-oriented, those employees, more than ever before, will need to be able to work cooperatively with others and interact effectively with the public.[16] Moreover, they will need to have completed high school and, for roughly three out of four new jobs, some education or training beyond high school.[17]

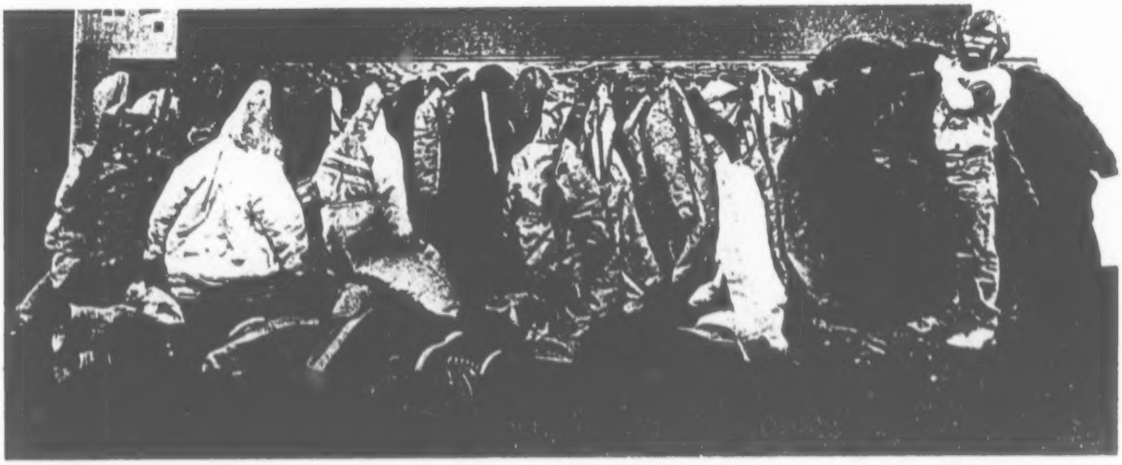
Our one hundred typical Maine youngsters will need to adapt quickly to change as the ever shorter life cycle of products, technologies, and industries results in more rapid obsolescence of their skills.[18] They will change jobs an average of eight times during their working lives.[19] As knowledge continues to multiply, information-management and computer-related skills will become tools for survival.

As we survey the current state of our delivery on that educational promise to Maine children, we must be alarmed. In the face of this rapid change, our public schools retain an outdated model of schooling at great costs to the young — and ultimately to society at large. The quality of education once reserved for a relative few must now be offered to all.

Most public high schools in Maine and the nation are, in fact, two high schools: one for students in college preparatory courses, the other for the rest. The tracking system, which sorts youngsters into at least these two categories (and frequently into more), purports to separate by academic ability. The evidence suggests that it actually separates by socio-economic status.

Beginning gradually with reading groups in the early grades, such grouping practices become less flexible as children grow older. As students in the "lower" track begin high school, they frequently become locked into a program that offers less adequate curriculum and instruction than that provided to their college-bound peers. Adults often underestimate the intellectual abilities of "lower track" students, sometimes by a wide margin, and either subtly or overtly communicate low expectations for academic performance. Inequalities in commitment to the two student groups abound, shown in distribution of resources, teaching skill, exploration of rich and stimulating ideas, adult enthusiasm, and positive messages about students' worth.

**The quality of education once reserved for a relative few must now be offered to all.**



School communities that wish to examine evidence of these inequalities should work with the available data about the relationship between socio-economic status and students' placement in academic tracks or "ability groups" in their own schools. (The argument that students choose their own placement is specious; by the time students reach high school, they have repeatedly received clear messages about where they "belong" in the tracking system.)

School communities might also examine their students' performances on the eleventh grade Maine Educational Assessment (MEA), designed to test not what college-bound juniors should know but what *every* Maine junior should know. Check the percentage of students enrolled in college-preparatory versus other programs. Look at the disparity between the average MEA scores of those two groups. Study the scores as broken down by gender. What conclusions can you reach about the degree to which our schools, as they are now structured, serve the needs of all students?

When surveyed by the MEA, teachers and principals across the state acknowledge that their schools' academic programs are less adequate in meeting the needs of non-college-bound students than the needs of the college-bound. More powerful evidence comes from students themselves. Concerned educators and citizens may listen to youngsters talk about the ways in which schools separate students and what that actually means for them. The Commission's student readers at three Maine high schools spoke eloquently to this, and our experiences in schools underscore their statements.

**Most public high schools ... are, in fact, two schools: one for students in college preparatory courses, the other for the rest.**

An alternative way to learn about these inequalities might be to spend one school day following the schedule of a non-college-bound student and the next in college preparatory classes. Adults who have done this describe significant differences between those experiences and relay great concern about the degree to which we are underserving about half the student population.

Too many of our students are already not being prepared to fulfill their potential, to have meaningful choices about their employment, and to meet the responsibilities of citizenship in a democratic society. All our youngsters are Maine's most precious resource. We must prepare all for the demands of their future.

With strong community support, our schools can adapt to the challenges of a rapidly-changing, information-based society. We can prepare students for the society in which they will live. The Common Core of Learning is a renewed promise to the youth of Maine. It is a commitment to an education that will offer the knowledge, skills, and attitudes they will need to ensure their independence, choice, and quality of life in the twenty-first century.

J'aime beaucoup
étudier
le français

**All our youngsters are Maine's most precious resource. We must prepare all for the demands of their future.**

# THE COMMON CORE OF LEARNING

*In the twenty-first century, the world will be our classroom and the universe shall be studied. — JASON BECKLER*[1]

As we discuss what our youth should know and be able to do in the twenty-first century, it is tempting to chip away at a very complex, integrated whole by relying on the familiar ways of dividing knowledge: the subject areas, as they were divided for us when we studied them in school. We are, however, entering an era in which we are thinking differently about knowledge — in fact, looking at a different conceptual model of how our world works. Observations of complexity and chaos both within the atom and among the reaches of the cosmos teach us that our search for the orderly, predictable arrangement of pieces results in only a partially accurate picture of reality. The same lesson comes out of studies of the inner workings of the human brain, which processes and organizes many things — parts and wholes — simultaneously. These models, emphasizing holism and integration, reshape not only the way we conceptualize the world around us, but also the way we think about learning.[2]

The knowledge, skills, and attitudes presented here are loosely organized into four categories that cut across the familiar subject areas. While we recognize that some thought processes are unique to a given subject (thinking like a historian, for instance, is not the same as thinking like a scientist), we also recognize that some areas are common to many subjects. Let's think of these areas in musical terms, as dynamic, complex, unified sound — an artistic whole.

We know that musicians are natural integrators: they wear many hats. They are linguists, communicating in the only language understood in every culture on the globe. They are physicists, adjusting nuances of their technique to their understanding of sound's travel through air in varying kinds of spaces. They are mathematicians, subdividing rhythms and pitches and using proportion in volume, phrasing, and tone color. They are physical and intellectual athletes, with highly disciplined muscles of body and sinews of memory. They are problem-solvers, alert, instantly flexible. They are psychologists in their understanding of a composer's intent and in their intuitive sense of what their fellow musicians are doing.

**These models, emphasizing holism and integration, reshape . . . the way we think about learning.**

Most of all, they are artists, engaged in dynamic recreation and expression of the deepest parts of the self — the essence of what it means to be human. All this at once!

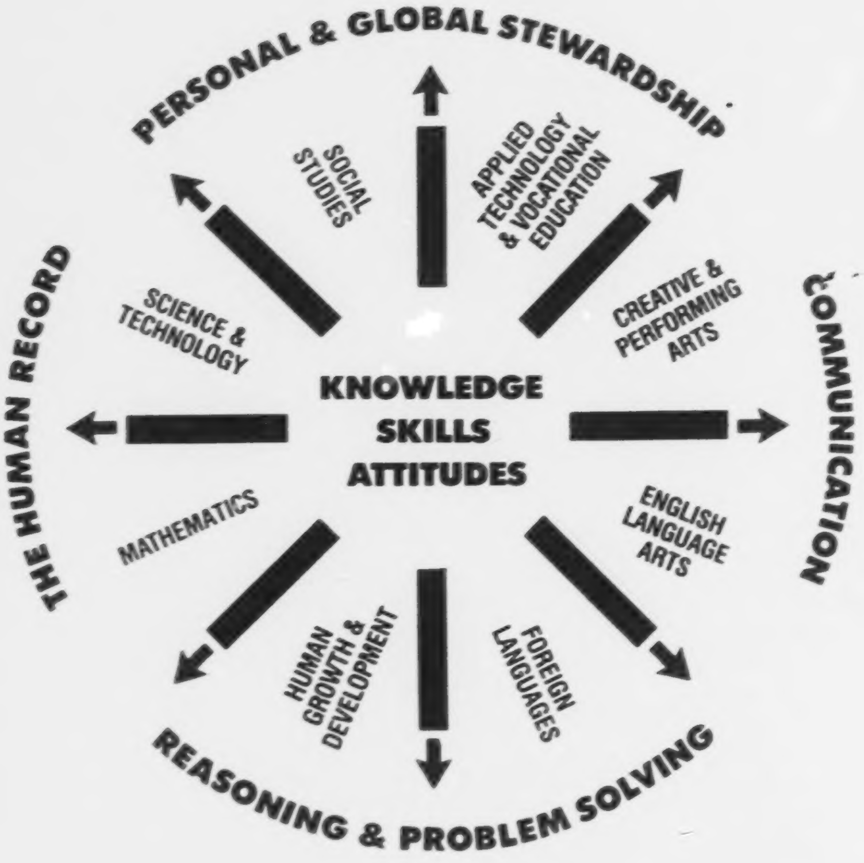
We, as listeners, hear their music all at once; it engages and touches us at many levels as we respond to its complexity.

Similarly, teaching and learning in the twenty-first century must have an integrated approach. Increasingly we will need to see connections and patterns among areas of thought that have too often been understood — and taught — in compartments or pieces. In our daily lives, as in making or hearing music, we do operate in an integrated or "interdisciplinary" fashion. Consider the knowledge, skills, and judgments involved in driving a car, preparing a meal, using your checkbook, or cultivating a garden.

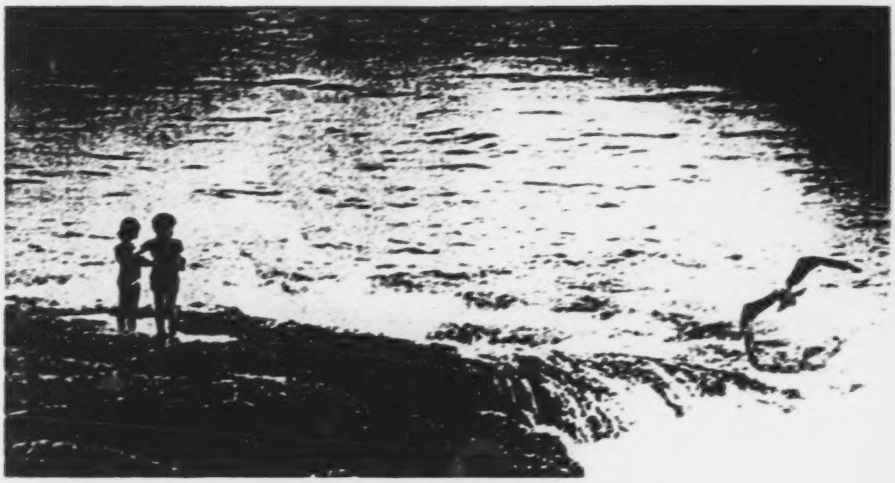
Good teaching has always helped students to see the world and their knowledge as a whole and to be able to think critically and flexibly. In the future, these abilities will be ever more important. Our educational process must move away from the fragmenting of knowledge that has characterized it in the past. This is no easy task — it involves reconstructing deeply-internalized patterns of thought — yet it is necessary if we want our students to think better.

The Common Core of Learning is a non-disciplinary organization of knowledge, skills, and attitudes, an integrated approach to teaching and learning. It is *not* an arrangement of skills into four new courses (Communication I, The Human Record II). The essential concepts gleaned from the currently taught subject areas are organized in the areas of Personal and Global Stewardship, Communication, Reasoning and Problem Solving, and The Human Record. They are listed by subject areas in the Appendix.

**Our educational process must move away from the fragmenting of knowledge that has characterized it in the past.**


PERSONAL & GLOBAL STEWARDSHIP
THE HUMAN RECORD
COMMUNICATION
REASONING & PROBLEM SOLVING
SOCIAL STUDIES
APPLIED TECHNOLOGY & VOCATIONAL EDUCATION
SCIENCE & TECHNOLOGY
CREATIVE & PERFORMING ARTS
KNOWLEDGE
SKILLS
ATTITUDES
MATHEMATICS
ENGLISH LANGUAGE ARTS
HUMAN GROWTH & DEVELOPMENT
FOREIGN LANGUAGES

THE COMMON CORE OF LEARNING



## PERSONAL AND GLOBAL STEWARDSHIP

Responsible citizenship requires awareness and a concern for oneself, others, and the environment. It involves interactions not only within the self and family, but between the self and friends, the community, the nation, and the world. It includes the knowledge and care of all dimensions of our selves as humans, an understanding of the group process, and a willingness to exercise the rights and responsibilities of citizenship. Stewardship also includes the study of current geography and foreign language and an appreciation of pluralism and human rights.

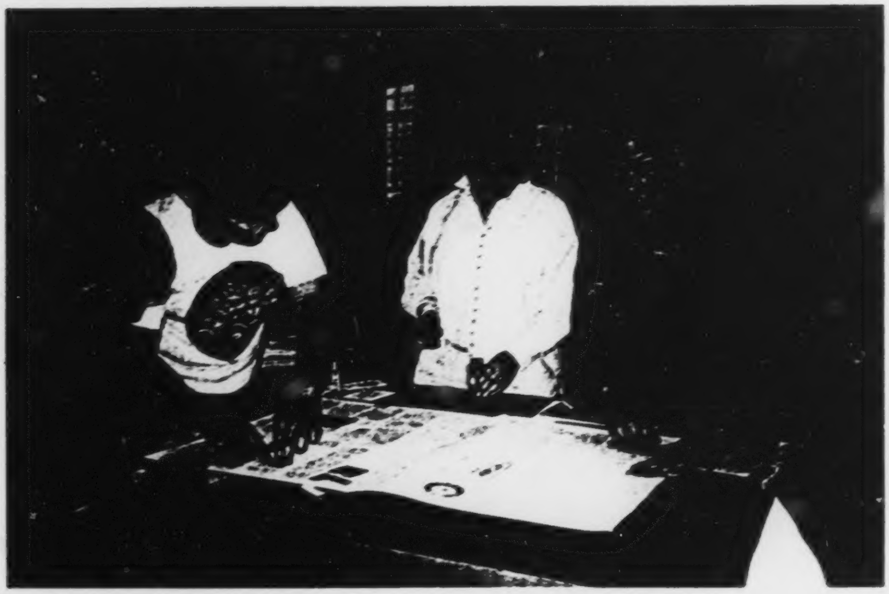
**Students with a common core of knowledge . . .**

- Accept responsibility for personal decisions and actions
- Have self-confidence and a willingness to risk mistakes in order to learn
- Are responsible for their own learning as independent and cooperative thinkers and informed decisionmakers
- Have the ability to adapt quickly to new situations and react to new information
- Concentrate and persevere on tasks
- Demonstrate academic honesty and respond to challenges with courage and integrity
- Respect the human rights of all people
- Understand the ethical dimensions of citizenship, love, friendship, and parenting
- Have skills that enhance their personal well-being:
  - Ability to state their own needs
  - Decision-making ability, including the ability to make informed career and life choices that result in healthy self-esteem

  - Straightforward communication and ability to build and sustain healthy relationships
  - Ability to manage stress
  - Ability to cope successfully with peer pressure and media messages urging dangerous or unhealthful behavior
  - Ability to select leisure time activities that develop and enhance wellness
- Understand basic concepts of growth and development, sexuality, family life, and parenting
- Have a basic understanding of the changing roles and rights of women and men in an increasingly diverse society
- Understand and practice self-care health skills, including good nutrition, safety and first aid, and avoidance of alcohol, tobacco, and other drugs
- Know when, where, and how to gain access to good health care
- Participate daily in physical activity and assess, develop, and maintain physical fitness
- Have a basic understanding of the development of modern technology and its effects on people, human culture, and the environment
- Find satisfaction in reading and writing and make those activities part of their everyday lives
- Understand that calculators, computers, and other technological devices are tools in the hand of the learner
- Believe that mathematics makes sense and have confidence in their ability to use mathematics meaningfully
- Show empathy and courtesy for others and respect for differences among people and cultures
- Understand the nature and roots of prejudice in themselves and others, as well as the ways prejudice contributes to injustice and oppression
- Are aware of our interdependence with the environment and are socially and environmentally responsible consumers
- Understand the arts to be a process for personal development and expression as well as for creating an aesthetic product
- Grasp the unique ability of the arts to encourage empathy and build a sense of community
- Possess knowledge of the everyday culture of another country (major customs, cultural commonplaces, and daily routines) and can identify similarities and differences between it and American culture
- Possess, as a result of non-native language study, an understanding of and an appreciation for the place of their own culture, language, and historic/ethnic heritage, as well as those of others, in a

pluralistic society

- Understand the geography of the places where a non-native language is spoken
- Understand basic principles of individual and group behavior, social organizations, and the processes of social change
- Understand the contributions made by various racial, ethnic, and religious groups to pluralistic societies, both their own and others around the world
- Demonstrate basic knowledge of the physical, economic, social, cultural, and political geography of Maine, the United States, and the world
- Understand the historical evolution of democratic principles and components of the constitutional government of the United States, as well as policy-making procedures of our local, state, and federal governments
- Show a basic understanding of world economic ideologies, systems, and practices and the global interdependence of economies
- Value and experience the opportunities for political, social, and economic participation in the life of the community
- Appreciate the inherent value of natural resources apart from their use by humans
- Respect and value all forms of life, their role in whole and healthy ecosystems, and the precarious position of our planet's environment in supporting life
- Understand work both as a means of economic survival and as an important source of personal identity and satisfaction
- Have a basic knowledge of work — careers, occupations, jobs, and the structure and functions of the labor market (national, state, and local)
- Are aware of their individual interests, aptitudes, skills, and values in relation to demands of the workplace; set goals for future work and understand the necessary preparation
- Understand the value of dependability, productivity, and initiative in all areas of life, including the world of work
- Seek out a fair share of the work load and manage time responsibly
- Work cooperatively and actively in group decision making, whether in small groups or in the larger society; are able to listen, share opinions, negotiate, compromise, and help the group reach consensus



## COMMUNICATION

The ability of human beings to communicate through a variety of media with a high degree of specificity is one of our most remarkable achievements. In a rapidly-changing world, communications skills will become ever more essential to our students' future success.

**Students with a common core of knowledge . . .**

- Communicate clearly — orally, in writing, and with graphics
  - Have a strong command of standard oral and written language conventions
  - Demonstrate basic proofreading and editing skills
  - Use handbooks and reference books to locate language terminology and rules
  - Use writing to record the thoughts of others
  - Reflect on and evaluate their own language use, including choosing the level of language (formal, colloquial, slang) appropriate to the setting in which they find themselves
  - Understand the role of symbols as a means of human expression
- Use oral and written language in all its varieties to get things done, to take charge of their lives, to express their opinions, to function as productive citizens, and to entertain and enjoy themselves and others
- Demonstrate awareness of gender stereotyping and cultural bias in our language

- Can use technology-based (CD ROM, interactive video, multimedia, telecommunication, etc.) and traditional ways of acquiring information
- Understand the similarities and differences in the ways various social, vocational, and cultural groups use language
- Can use
    - Database management skills to organize information
    - Word-processing skills to convey ideas
    - Spreadsheet and graphing skills to express ideas quantitatively
- Understand the capabilities of the computer and use it as a creative learning tool
- Can express their ideas and emotions through participation in one or more of the visual and performing arts
- Can use and understand language appropriate to each art form when discussing and interpreting art works
- Can communicate and understand others in more than one language
    - Can ask and answer questions in another language and maintain a simple conversation in areas of immediate personal need
    - Can read, understand, and write in individual sentences or a short paragraph in another language authentic information about everyday topics
- Can understand and use the language of mathematics, orally and in writing, including the ability to express mathematical ideas through measurement and with physical material, pictures, diagrams, and graphs
- Have sufficient knowledge of the vocabulary and methods of science to be intelligent consumers and responsible users of scientific information
- Have mastered the numerical skills of science, including basic dexterity with numbers, competency with simple calculators, the ability to make estimations, and an appreciation for the vast differences among orders of magnitude from the subatomic to interstellar space

## REASONING AND PROBLEM SOLVING

Knowledge is power. We must help students want to gain knowledge, show them how to get it, and encourage them to use it to reach a new understanding or to create a new product. We must help students learn to reflect on their processes of learning, regardless of their field of study.

**Students with a common core of knowledge . . .**

- Have developed attitudes and thinking skills essential to lifelong learning: curiosity and openness to new ideas, creativity, integrity, diligence, fairness, skepticism, and imagination
- Can think logically and solve problems by
    - Observing carefully
    - Defining the problem
    - Framing useful questions
    - Designing experiments that can answer those questions
    - Gathering and selecting information
    - Analyzing data — comparing and contrasting, seeking patterns, deducing sequential order, and developing tables/graphs/rules
    - Considering and testing more than one solution
    - Justifying strategies and solutions
    - Applying strategies and solutions to new situations
- Find tools in all areas of study to solve problems, investigate content, and develop thinking skills

- Recognize the interrelationships of the basic principles within an area of study
- Can easily make connections and move between and among the various areas of study
- Use language to understand themselves and others, to make sense of their world, and to reflect on their lives — by telling and listening to stories; by reading novels, poetry, and biographies; and by keeping journals or logs
- Can, as competent language users,
    - Use prior knowledge to comprehend new oral or written texts
    - Know when they don't know or don't understand oral or written language and have strategies for clarifying their thinking — by asking questions, rereading, or writing to learn
    - Respond to the material they read, hear, or watch by questioning, connecting, evaluating, and extending
    - Understand literal messages read or heard
    - Read and listen critically and interpretively
    - Make connections within and among texts
    - Use other readers' experiences with, responses to, and interpretation of texts
    - Hear literature, appreciating its sounds and cadences
- Recognize when language is being used to manipulate, coerce, or control them, and use language as an effective response to such attempts
- Understand that writing is a process that involves planning, drafting, giving and receiving feedback, revising, editing, and publishing
- Sustain concentration and commitment to problem solving over necessary lengths of time using oral or written language as a means of organizing and recording thoughts
- Understand enough about their individual learning styles to explain how they best become able to talk and/or write about an idea
- Understand how the arts, because of their power to move us, can be misused to exploit and manipulate masses of people
- Can create an original work in one of the fine arts or perform an interpretation of an existing work in one of the performing arts
- Can use the formal elements of the arts to make aesthetic judgments
- Can make number sense
    - Use estimation in both solving problems and in checking the reasonableness of the results
    - Explore relationships among representations of, and operations on, whole numbers, fractions, decimals, integers, and rational numbers

- Select and use an appropriate computational method from among mental arithmetic, paper-and-pencil, calculator, and computer methods
- Understand and can apply concepts of ratio, proportion, and percent in a variety of situations
- Use mathematical models, facts, properties, and relationships to explain their thinking
- Can create experimental and theoretical models of situations involving probabilities using traditional and computer-based technologies
- Use basic statistical methods to describe, analyze, evaluate, and make decisions
- Can represent a situation that involves variable quantities with expressions, equations, inequalities, and matrices (basic abstract algebra)
- Recognize and formulate real-world problems from situations within and outside mathematics
- Understand personal economics and have the ability to manage money
- Demonstrate fundamental scientific manipulation skills — reading scales and gauges, diluting liquids, mixing solid substances based on prescribed proportions, trouble-shooting common mechanical problems, and making simple electrical connections

## THE HUMAN RECORD

The study of the human record not only includes the actions and events of the past but also the constructs of human thought and creativity as they have evolved through time. The human record includes works of literature and the arts; scientific laws and theories; and concepts of government, economic systems, philosophy, and mathematics. In fact, much of what we now think of as "subject matter" in today's curriculum belongs in this section.

**Students with a common core of knowledge . . .**

- Are able to read, for a variety of purposes, diverse types of material, including poems, stories, novels, drama, and expository texts in both print and electronic media
- Are familiar with works of diverse literary traditions — works by women and men of many racial, ethnic, and cultural groups in different times and parts of the world, including Shakespeare, the Bible as literature, and classical mythology
- Are familiar with contemporary and enduring works of American literature and have a sense of how important themes of American experience have developed through time
- Have a basic understanding of the history and structure of the English language
- Understand and appreciate the unique qualities of each of the art forms — drama, dance, music, and the visual arts
- Have a basic knowledge and understanding of art forms from various cultures and historical periods

- Understand how libraries and other repositories of human knowledge are organized
- Know about themselves as a species — the human life cycle, how we are distinguished from other life forms, the ways we learn and reason, and how we function biologically
- Know how to prevent or control diseases and disorders, including HIV (AIDS) infection
- Understand the role of daily physical activity in contributing to personal health and well-being
- Realize that scientific theories are creations of the human mind and can change with new developments
- Appreciate the impact of significant scientific discoveries and grasp the changes in human history brought about by the evolution of scientific thought
- Understand the atomic basis of the structure of matter, the existence of the basic forces of nature, the behaviors of the various stages of matter, and the common changes that various types of matter undergo
- Are aware of the forces that shape the earth as well as the universe and recognize that both order and chaos are present
- Appreciate that transformations of energy pervade biological, chemical, geological, and physical systems
- Understand the basic mechanical, optical, electrical, and magnetic workings of the physical world as well as the fundamental laws governing the motion of objects
- Understand the roles played by cell theory, plant and animal structure and function, ecological relationships, the principles of genetics, and evolution in our living environment
- Have a working understanding of the concepts, processes, and systems of technology, and of the safe and effective application of tools, materials, and machines
- Understand and apply the core concepts embedded in each of the social sciences — history, geography, political science, economics, philosophy, sociology, psychology, and the humanities
- Understand and be able to apply fundamental themes in geography: location, place, relationships within places, movement, regions, and global interdependence
- Understand the eras and major concepts in world history and in American history
- Understand the democratic principles upon which the United States was founded
- Demonstrate a working knowledge of the Declaration of Independence and the United States Constitution

# III-12

# National Education Goals and Objectives

U.S. Congress

The Congress declares that the National Goals are the following:

## SCHOOL READINESS

**By the year 2000, all children in America will start school ready to learn.**

The objectives for this goal are that

- all children will have access to high-quality and developmentally appropriate preschool programs that help prepare children for school;
- every parent in the United States will be a child's first teacher and devote time each day to helping such parent's preschool child learn, and parents will have access to the training and support parents need; and
- children will receive the nutrition, physical activity experiences, and health care needed to arrive at school with healthy minds and bodies, and to maintain the mental alertness necessary to be prepared to learn, and the number of low-birthweight babies will be significantly reduced through enhanced prenatal health systems.

## SCHOOL COMPLETION

**By the year 2000, the high school graduation rate will increase to at least 90 percent.**

The objectives for this goal are that

- the nation must dramatically reduce its school dropout rate, and 75 percent of the students who do drop out will successfully complete a high school degree or its equivalent; and
- the gap in high school graduation rates between American students from minority backgrounds and their non-minority counterparts will be eliminated.

## STUDENT ACHIEVEMENT AND CITIZENSHIP

**By the year 2000, all students will leave grades 4, 8, and 12 having demonstrated competency over challenging subject matter including English, mathematics, science, foreign languages, civics and government, economics, arts, history, and geography, and every school in American will ensure that all students learn to use their minds well, so they may be prepared for responsible citizenship, further learning, and productive employment in our nation's modern economy.**

The objectives for this goal are that

- the academic performance of all students at the elementary and secondary level will increase significantly in every quartile, and the distribution of mi-

---

Reprinted from Goals 2000: Educate America Act (PL 103-227). Update: April 1994.

nority students in each quartile will more closely reflect the student population as a whole;

- the percentage of all students who demonstrate the ability to reason, solve problems, apply knowledge, and write and communicate effectively will increase substantially;
- all students will be involved in activities that promote and demonstrate good citizenship, good health, community service, and personal responsibility;
- all students will have access to physical education and health education to ensure they are healthy and fit;
- the percentage of all students who are competent in more than one language will substantially increase; and
- all students will be knowledgeable about the diverse cultural heritage of this nation and about the world community.

### TEACHER EDUCATION AND PROFESSIONAL DEVELOPMENT

**By the year 2000, the nation's teaching force will have access to programs for the continued improvement of their professional skills and the opportunity to acquire the knowledge and skills needed to instruct and prepare all American students for the next century.**

The objectives for this goal are that

- all teachers will have access to preservice teacher education and continuing professional development activities that will provide such teachers with the knowledge and skills needed to teach to an increasingly diverse student population with a variety of educational, social, and health needs;
- all teachers will have continuing opportunities to acquire additional knowledge and skills needed to teach challenging subject matter and to use emerging new methods, forms of assessment, and technologies;
- states and school districts will create integrated strategies to attract, recruit, prepare, retrain, and support the continued professional development of teachers, administrators, and other educators, so that there is a highly talented work force of professional educators to teach challenging subject matter; and
- partnerships will be established, whenever possible, among local educational agencies, institutions of higher education, parents, and local labor, business, and professional associations to provide and support programs for the professional development of educators.

### MATHEMATICS AND SCIENCE

**By the year 2000, United States students will be first in the world in mathematics and science achievement.**

The objectives for this goal are that

- mathematics and science education, including the metric system of measurement, will be strengthened throughout the system, especially in the early grades;
- the number of teachers with a substantive background in mathematics and science, including the metric system of measurement, will increase by 50 percent; and
- the number of United States undergraduate and graduate students, especially women and minorities, who complete degrees in mathematics, science, and engineering will increase significantly.

### ADULT LITERACY AND LIFELONG LEARNING

**By the year 2000, every adult in America will be literate and will possess the knowledge and skills necessary to compete in a global economy and exercise the rights and responsibilities of citizenship.**

The objectives for this goal are that

- every major American business will be involved in strengthening the connection between education and work;
- all workers will have the opportunity to acquire the knowledge and skills, from basic to highly technical, needed to adapt to emerging new technologies, work methods, and markets through public and private educational, vocational, technical, workplace, or other programs;
- the number of quality programs, including those at libraries, that are designed to serve more effectively the needs of the growing number of part-time and midcareer students will increase substantially;
- the proportion of the qualified students, especially minorities, who enter college, who complete at least two years, and who complete their degree programs will increase substantially;
- the proportion of college graduates who demonstrate an advanced ability to think critically, communicate effectively, and solve problems will increase substantially; and
- schools, in implementing comprehensive parent involvement programs, will offer more adult literacy, parent training and lifelong learning opportunities to improve the ties between home and school, and enhance parents' work and home lives.

## SAFE, DISCIPLINED, AND ALCOHOL-AND DRUG-FREE SCHOOLS

**By the year 2000, every school in the United States will be free of drugs, violence, and the unauthorized presence of firearms and alcohol and will offer a disciplined environment conducive to learning.**

The objectives for this goal are that

- every school will implement a firm and fair policy on use, possession, and distribution of drugs and alcohol;
- parents, businesses, governmental and community organizations will work together to ensure the rights of students to study in a safe and secure environment that is free of drugs and crime, and that schools provide a healthy environment and are a safe have for all children;
- every local educational agency will develop and implement a policy to ensure that all schools are free of violence and the unauthorized presence of weapons;
- every local educational agency will develop a sequential, comprehensive kindergarten through twelfth grade drug and alcohol prevention education program;
- drug and alcohol curriculum should be taught as an integral part of sequential, comprehensive health education;
- community-based teams should be organized to provide students and teachers with needed support; and
- every school should work to eliminate sexual harassment.

## PARENTAL PARTICIPATION

**By the year 2000, every school will promote partnerships that will increase parental involvement and participation in promoting the social, emotional, and academic growth of children.**

The objectives for this goal are that

- every state will develop policies to assist local schools and local educational agencies to establish programs for increasing partnerships that respond to the varying needs of parents and the home, including parents of children who are disadvantaged or bilingual, or parents of children with disabilities;
- every school will actively engage parents and families in a partnership which supports the academic work of children at home and shared educational decision making at school; and
- parents and families will help to ensure that schools are adequately supported and will hold schools and teachers to high standards of accountability.

# III-13

# Oregon's Certificate of Initial Mastery Outcomes and Extended Outcome Definitions (Draft)

OREGON DEPARTMENT OF EDUCATION

Reprinted with permission of the Oregon Department of Education, Salem. Contributed by Carol Black, Education Specialist, School Reform.

208

DRAFT

## Student Outcomes

To attain the Certificate of Initial Mastery, a student will demonstrate the ability to:

### Core Applications for Living

| | |
|---|---|
| **Deliberate on Public Issues** | deliberate on public issues which arise in our representative democracy and in the world by applying perspectives from the social sciences |
| **Understand Diversity** | understand human diversity and communicate in a second language, applying appropriate cultural norms |
| **Interpret Human Experience** | interpret human experience through literature and the visual and performing arts |
| **Apply Science and Math** | apply science and math concepts and processes, showing an understanding of how they affect our world |
| **Understand Positive Health Habits** | understand positive health habits and behaviors that establish and maintain personal wellness and healthy interpersonal relationships |

### Foundation Skills

| | |
|---|---|
| **Think** | think critically, creatively and reflectively in making decisions and solving problems |
| **Self-Direct Learning** | direct his or her own learning, including planning and carrying out complex projects |
| **Communicate** | communicate through reading, writing, speaking, and listening, and through an integrated use of visual forms such as symbols and graphic images |
| **Use Technology** | use current technology, including computers, to process information and produce high quality products |
| **Quantify** | recognize, process, and communicate quantitative relationships |
| **Collaborate** | participate as a member of a team, including providing leadership for achieving goals and working well with others from diverse backgrounds |

# Extended Outcome Definitions

The Certificate of Initial Mastery (CIM) outcomes and the extended outcome definitions have been adopted by the State Board of Education. The CIM outcomes emphasize useful knowledge and complex performances, matched to real-world demands. The outcomes are relatively few in number and provide a focused, uncluttered set of targets for learning.

The extended outcome definitions presented on the following pages provide a more detailed description of the CIM outcomes. The extended outcome definitions are not intended to identify curriculum goals, describe developmentally appropriate activities, or define expected levels of performance. However, because the extended definitions do illustrate mental pictures of student performances, they will prove useful to those who develop program, assessment tasks and scoring criteria for the Certificate of Initial Mastery.

To measure student progress relative to the CIM outcomes, additional components of the CIM assessment system are being developed. These components include scoring scales (rubrics), model assessment tasks and expected performance levels. Once all the components of the CIM assessment system are developed and can be understood in relation to each other, the CIM standards for student performance will be defined.

DRAFT

## Core Application: Deliberate on Public Issues

### DELIBERATE ON PUBLIC ISSUES

| | |
|---|---|
| CIM OUTCOME: | Deliberate on public issues which arise in our representative democracy and in the world by applying perspectives from the social sciences. |
| Extended definition: | Students will propose relevant actions for local, state, national and international public issues. A student who effectively deliberates on public issues is able to:<br>• identify and clarify relevant issues<br>• analyze issues systematically (e.g., identify resources, analyze diverse perspectives, make predictions and form conclusions)<br>• apply democratic principles such as fairness and justice in formulating options and actions<br>• apply information, concepts and perspectives from history, the social sciences and other relevant disciplines |

DRAFT

# Core Application: Understand Diversity
# Communicate in a Second Language

## UNDERSTAND DIVERSITY

**CIM OUTCOME:**

Understand diversity
Communicate in a second language, applying cultural norms, as appropriate.

**Extended definition:**

(1) understand diversity

Students will respond to diversity in a way that recognizes the dignity and rights of all people. A student who understands diversity is able to:

- analyze his or her own culture, recognizing the influences that have shaped thinking and behavior
- explain how our perceptions of differences and similarities among people (e.g., cultural, ethnic, racial, ability level, gender) enrich our lives or may lead to stereotyping, miscommunication, discrimination and the denial of human rights
- analyze systematically the interaction between cultural populations, in order to gain awareness and sensitivity to both human diversity and cultural identity
- recommend strategies to enhance communication and collaboration, reduce tensions, and resolve misperceptions and conflicts relating to human diversity

**Extended definition:**

(2) communicate in a second language*

Students will be able to communicate with a person from another culture in that person's language and show a sensitivity to the norms of behavior that apply. A student who can communicate in a second language is able to:

- comprehend and respond to spoken messages and commands
- maintain simple conversation that shares information
- recognize and show sensitivity to body language, gestures and appropriate levels of formality and other aspects of the culture which affect communication
- read and comprehend basic material encountered in everyday life
- write to meet practical needs

*In instructing students, emphasis will be given to oral communication.

If the student's native language is other than English, the student may be tested for proficiency in this outcome in that language.

DRAFT

## Core Application: Interpret Human Experience

### INTERPRET HUMAN EXPERIENCE

| | |
|---|---|
| CIM OUTCOME: | Interpret human experience through literature and the visual and performing arts. |
| Extended definition: | Students will relate literature and the arts to their own lives and to broader human concerns, issues and possibilities. A student who derives meaning and value from literature and the arts is able to:<br>• make informed interpretations of literary and artistic works<br>• create, improvise, or perform works of art (e.g., literature, visual arts, music, drama, dance)<br>• explain how literature and the arts from various cultural/ethnic groups express distinctive as well as common values, experiences, struggles and contributions<br>• explain how the form and content of a literary or artistic work contribute to its aesthetic impact |

DRAFT

## Core Application: Apply Science and Mathematics

### APPLY SCIENCE AND MATHEMATICS

CIM OUTCOME: Apply science and math concepts and processes, showing an understanding of how they affect the world.

Extended definition: Students will apply mathematics and science to varying real world situations. A student who is proficient in these applications is able to:

- analyze real world phenomena using scientific concepts (e.g., cause and effect, energy, systems)
- apply mathematics, scientific laws and processes, and technological systems to real-world problems
- construct mathematical and scientific models
- analyze the impact of technology on economic, social, political and environmental systems

DRAFT

## Core Application: Understand Positive Health Habits

### UNDERSTAND POSITIVE HEALTH HABITS

| | |
|---|---|
| CIM OUTCOME: | Understand positive health habits and behaviors that establish and maintain personal wellness and healthy interpersonal relationships. |
| Extended definition: | Students will assess and monitor in their own lives the factors that relate to positive mental, physical and social health. A student who understands the factors which contribute to the development of a healthy individual is able to:<br>• analyze the relationships among health factors, disease prevention and health promotion<br>• assess personal health risks and apply this assessment to develop methods to reduce the risks<br>• predict own future health status based on current health habits and knowledge of the life cycle<br>• recognize conditions, actions and personal qualities which affect interpersonal relationships in the specific life roles of citizen, family member and worker |

DRAFT

## Foundation Skill: Think

### THINK

**CIM OUTCOME:** Think critically, creatively and reflectively in making decisions and solving problems.

**Extended definition:**

**(1) think critically**

Students will identify and evaluate reasons, assumptions and evidence that support their own positions or practices or the positions and practices of others. A student who thinks critically is able to:

- define and clarify problems
- determine accuracy and relevancy of information
- use language precisely to make distinctions
- select effective strategies and procedures to solve problems and investigate questions
- construct sound inferences, defensible conclusions, and meaningful interpretations
- make connections between prior knowledge and experience and new issues, problems, and learnings
- examine situations/problems from multiple perspectives
- use metaphoric comparisons and analogies, when appropriate, to clarify the essence of a problem and to open up new ideas for its solution
- make judgments and evaluations based on principles or criteria

**(2) think creatively**

Students will think flexibly and insightfully, going beyond routine or superficial responses when framing problems and seeking solutions, developing plans and designing products and performances, and engaging in other complex tasks. A student who thinks creatively is able to:

- create meaning from unstructured experience and information using such processes as recognizing patterns, developing concepts and generalizations, and posing questions and hypotheses
- generate new ideas or combine existing ideas and viewpoints in new ways, to produce meaningful and effective alternatives, plans, proposals, and solutions
- see new relationships and develop symbols to represent those relationships
- develop interpretations and performances that reflect individuality, independent judgment, and inventiveness

**(3) think reflectively**

Students will consciously monitor, assess and improve their own thinking. A student who thinks reflectively is able to:

- identify thinking strategies (for example, deliberating, problem solving, investigating, evaluating, synthesizing) that are appropriate for various situations and select an approach that is likely to be successful
- assess the strengths and weaknesses of the frames of reference (for example, historical, scientific, cultural, ethical, legal, political, metaphoric) being applied in various situations and select frames that are likely to be most helpful in accomplishing one's purpose
- analyze and modify thinking strategies to increase the chance of being successful at a task
- assess the accuracy and adequacy of the results and the appropriateness/effectiveness of the thinking strategies used

DRAFT

## Foundation Skill: Self-Direct Learning

### SELF-DIRECT LEARNING

CIM OUTCOME: Direct individual learning, including planning and carrying out complex projects.

Extended definition: Students will recognize and demonstrate their own power and capability to learn and solve problems. They will demonstrate that capability through the establishment of courses of action (including the planning and carrying out of complex projects) which have multiple steps, are open-ended, are challenging in scope and sophistication, involve research and the use of outside resources, and involve communication with other people.

**(1) direct individual learning**

A student who recognizes and demonstrates capability to learn and solve problems will be able to:

- understand the self as learner, including interests, aptitudes, abilities, and educational and career aspirations
- identify and access the resources necessary to obtain needed skills or knowledge
- establish clear goals and high standards for personal performance, be accountable to these standards, and persevere when faced with difficult situations
- analyze different learning environments and adapt strategies to improve learning

**(2) carry out complex projects**

A student who carries out complex projects is able to:

- set appropriate project goals and develop a well-reasoned plan for attaining the goals
- implement the plan, checking and adjusting activities to keep the project on track and persisting in the face of difficulties
- evaluate the final product(s) and reflect on the project as a whole, including what has been learned as a result

DRAFT

## Foundation Skill: Communicate

### COMMUNICATE

CIM OUTCOME: Communicate through reading, writing, speaking and listening and through an integrated use of visual forms such as symbols and graphic images.

Extended definition:

**(1) read**

Students will obtain meaning from a variety of complex texts (e.g., essays, novels, stories, poems, technical documents) and identify the writer's purposes. An effective reader is able to:

- make connections within and among texts and from personal experience
- choose reading strategies most appropriate to a given reading situation
- evaluate and monitor comprehension, using a variety of strategies to self-correct

**(2) write**

Students will use writing as a tool for learning and self-reflection and to convey meaning through a variety of written forms (e.g., stories, essays, journals, technical reports). An effective writer is able to:

- demonstrate control of ideas and content, organization, voice, word choice, sentence fluency and conventions
- evaluate and monitor personal writing, using a variety of strategies to produce coherent and mechanically correct final products

**(3) speak**

Students will engage critically and constructively in oral exchanges of information, including making formal presentations, giving spoken instructions, asking and answering questions, and using language to achieve effective group communication. An effective speaker is able to:

- deliver presentations, demonstrating effective skills relevant to the audience (e.g., vocal expression and non-verbal signals, development of ideas, organization, and level of language)
- communicate ideas effectively in group situations

**(4) listen**

Students will obtain meaning through oral messages, including information presented through a variety of media. An effective listener is able to:

- identify the purpose of an oral message
- analyze and evaluate verbal and non-verbal messages and how those messages are delivered
- engage in verbal and non-verbal interaction with a speaker to ensure effective communication

**(5) Integrate use of visual forms**

Students at the CIM level will select and develop visual forms to integrate with other forms of communication to enhance the impact of a product or presentation. A student who can integrate the use of visual forms is able to:

- integrate visual forms that achieve the intended purpose and are appropriate to the intended audience
- select and develop visual forms that meet the criteria for effective design (e.g., clarity, coherence, accuracy, precision)
- use visual forms that reflect efficient or creative use of available materials and technology

DRAFT

## Foundation Skill: Use Technology

### USE TECHNOLOGY

**CIM OUTCOME:** Use current technology, including computers, to process information and produce high quality products.

**Extended definition:** Students will use, develop and evaluate technology to gather, understand and manipulate materials, processes and information to solve problems. A technologically competent student is able to:

- apply diverse technologies to store, access, process, create and communicate information needed to solve problems, satisfy personal needs and extend human capabilities
- set up and maintain tools and equipment to produce high quality products
- use a variety of tools, materials and equipment safely

DRAFT

## Foundation Skill: Quantify

### QUANTIFY

| | |
|---|---|
| CIM OUTCOME: | Recognize, process and communicate quantitative relationships. |
| Extended definition: | Students will address a variety of challenges that require quantitative solutions (e.g., problems requiring the application of measurement, statistics, probability, geometry and algebra). A student who uses quantitative skills effectively is able to:<br>• interpret a problem situation, including selecting information to solve the problem<br>• develop and apply a problem solving strategy<br>• solve the problem and verify the solution<br>• communicate the results in an easily understood manner |

DRAFT

## Foundation Skill: Collaborate

### COLLABORATE

| | |
|---|---|
| CIM OUTCOME: | Participate as a member of a team, including providing leadership for achieving goals and working well with others from diverse backgrounds. |
| Extended definition: | Students will organize and actively engage in work groups aimed at achieving specific goals. An effective team member is able to:<br>• define problems, issues, strategies and tasks within the group<br>• perform the functions of various roles within the group<br>• establish and participate in open and clear communication<br>• engage all team members in efforts to achieve team goals<br>• work toward consensus while respecting divergent points of view |

DRAFT

# III-14

# Teaching a Core of Common Knowledge

TASK FORCE ON EDUCATION OF YOUNG ADOLESCENTS, CARNEGIE COUNCIL ON ADOLESCENT DEVELOPMENT

*Every student in the middle grades should learn to think critically through mastery of an appropriate body of knowledge, lead a healthy life, behave ethically and lawfully, and assume the responsibilities of citizenship in a pluralistic society.*

Every middle grade school should offer a core program and should expect every student to complete that program successfully. The purpose of setting this goal is practical: The future of this nation as a stable, prosperous democracy requires that all members contribute to the common good of society and meet their obligations as citizens.

The broad outlines of the curriculum are the responsibility of state and local school authorities. But middle grade educators can fashion a full academic program for all students that integrates English, fine arts, foreign languages, history, literature and grammar, mathematics, science, and social studies.

A curriculum containing these subjects today frequently lacks depth. At best, students am asked to demonstrate competency in the subject matter, but not in the ability to think or express themselves about topics that range across each subject. Middle grade schools, in conjunction with state and local authorities, can vastly improve middle grade curricula and instruction programs by:

- Teaching young adolescents to think critically;
- Teaching young adolescents to develop healthful lifestyles;
- Teaching young adolescents to be active citizens;
- Integrating subject matter across disciplines; and
- Teaching students to learn as well as to test successfully.

## Teaching Young Adolescents to Think Critically

Many middle grade schools in this country fail to support and challenge youth. Nowhere is this failure more evident than in the development of American young adolescents' critical reasoning and higher order thinking.*

Contrary to much conventional belief, cognitive development during early adolescence is not on hold. Belief in such claims has had substantial and damaging effects on middle grade education, by limiting innovation in curriculum development that might require new and more advanced ways of think-

---

Reprinted with permission from *Turning Points: Preparing American Youth for the 21st Century. The Report of the Task Force on Education of Young Adolescents* (Washington, D.C.: Carnegie Council on Adolescent Development, 1989), pp. 42-49. Copyright © 1989 by the Carnegie Council on Adolescent Development, Washington, D.C.

ing. A thorough review of recent studies on adolescent cognitive development found "no persuasive evidence" that young adolescents cannot engage in critical and higher order thinking.[2]

A primary goal in choosing curricula and teaching methods in the middle grades should be the disciplining of young adolescents' minds, that is, their capacity for active, engaged thinking. A student with a disciplined mind can assimilate knowledge, rather than merely recall information by rote or apply simple algorithms. The student can challenge the reliability of evidence; recognize the viewpoint or voice behind the words, pictures, or ideas presented; see relationships between ideas; and ask what-if and suppose-that questions.

A disciplined mind connotes a disposition toward inquiry, discovery, and reasoning across all subjects. A student cannot, of course, develop a disciplined mind apart from specific subject matter. The key lies in how the student approaches the subject matter. In the traditional curriculum, the student learns subject by subject. This fragmented array does not allow students to connect new and old ideas or to construct their own meaning of the information.

In the core curriculum of the transformed middle grade school, the student confronts themes, which are clusters of subjects, and learns to inquire, associate, and synthesize across subjects. The student learns to reason even while absorbing basic information about the subject matter. This approach clearly requires that the current emphasis on coverage of a large quantity of information must yield to an emphasis on depth or quality of the student's understanding. Schools can choose the most important principles and concepts within each subject, and concentrate instruction there.

Developing a discipline of mind requires a radical departure from traditional classroom instruction. Above all, it requires a shift in long-held beliefs that the role of the teacher is to transmit knowledge to students. Instead, teachers must view themselves as facilitators through which young people construct knowledge themselves. Teachers will need the support of administrators in their new role.

Teachers will be called upon to promote a spirit of inquiry and to stimulate students to think about and communicate ideas. Far greater reliance will be placed on learning techniques that allow students to participate actively in discovering and creating new solutions to problems.

For example, young adolescents need group approaches to learning. Learning often takes place best when students have opportunities to discuss, analyze, express opinions, and receive feedback from peers.[3] Peer involvement is especially critical during early adolescence when the influence of peers increases and becomes more important to the young person.

## Teaching Young Adolescents to Develop Healthful Lifestyles

Educators of young adolescents need to reconsider the contemporary approach to health education, which treats health education as a requirement of minor importance. Teachers of health education are too often not trained in the subject and devote little time to it.

One solution in middle grades is the integration of health education into the core instructional program as an element of the life sciences. Students who learn about health within the context of the life sciences find out about how their bodies grow and function. They come to realize the value of a healthful diet and exercise while recognizing the dangers of illicit drugs, alcohol, and tobacco.

A premier example of the integration of health and the life sciences can be found in the Human Biology Program at Stanford University. This highly successful interdisciplinary undergraduate major integrates the biological and behavioral sciences. Stanford educators are now adapting it from postsecondary education to the middle grades.

Life sciences education alone is unlikely to persuade young people to reject self-destructive or dangerous behavior. Young people need training in skills that help them to resist interpersonal or media messages to engage in specific negative behaviors, increase self-control and self-esteem, reduce stress and anxiety, gain in the ability to express apprehension and disapproval, and become assertive. These abilities are often referred to as life skills.

Students can learn these essential life skills through systematic instruction and role-playing. Efforts to teach adolescents these skills have been effective in preventing adolescent smoking[4] and substance abuse.[5] An intervention based on life skill development has also been effective in preventing pregnancies among older adolescents,[6] although its effectiveness in preventing sexual activity and pregnancy among young adolescents has not yet been evaluated.

As young people learn about their health and how to avoid problem behaviors, they must also gain access to health care and be able to exercise options in selecting care, important issues discussed later in this report.

## Teaching Young Adolescents to Be Active Citizens

Early adolescence offers a superb opportunity to learn values, skills, and a sense of social responsibility important for citizenship in the United States. Every middle grade school should include youth service—

supervised activity helping others in the community or in school—in its core instructional program-

Youth service can teach young people about values for full participation in this society. Educators agree that their mission includes teaching values for citizenship. These values certainly include compassion, regard for human worth and dignity, respect for others' rights, tolerance and appreciation of human diversity, willingness to cooperate, and a desire for social justice.

Youth service also teaches students about collaboration, problem solving, conflict resolution, and seeing projects in their entirety rather than in pieces. These skills are essential in the workplace.[7] Consequently, middle grade schools should look upon youth service as central to the academic program..

Service opportunities abound within the middle grade school itself. Students can serve as peer and cross-age tutors, teaching assistants, lab assistants, safety officers, school governance committee members, student court officers, assistants in coordinating youth service, and in many other roles. Youth service can expand beyond the boundaries of the school. Students who assist in childcare, help the handicapped, clean up a debris-strewn stream, or advocate for the homeless are assuming meaningful roles and responding to real problems. Integrating service into the core curriculum enables teachers to tie classroom instruction to real-life, hands-on learning experiences within the community.

Fully integrating youth service into the core program will require new roles for teachers and administrators to establish and coordinate youth-service opportunities. Teachers will need to provide opportunities for students to reflect on and learn from their service experience. Additional personnel, community service coordinators, may be needed to arrange student placements, coordinate with site supervisors, and oversee students' service activities on campus. The community can be expected to respond affirmatively to youth service, which offers many opportunities for partnerships with community organizations. Many youth organizations have long traditions of involving young people in service, both in the community and at their agencies, and are natural partners with middle grade schools for such a program.

In concert with the youth service program, schools must reflect the values they teach. An ethic of mutual respect must pervade the school, in the way school staff treat students and in the way staff treat each other. Appreciation of cultural diversity, application of equal opportunity practices, and sensitivity to ethnic and racial pluralism within the community should be reflected in the composition of the school staff, which in many schools will require the hiring of more minority teachers.

## Integrating Subject Matter Across Disciplines

Young adolescents demonstrate an ability to grapple with complexity, think critically, and deal with information as parts of systems rather than as isolated, disconnected facts. Currently, middle grade curricula package ideas and information strictly by subject, making it difficult for students to see connections between ideas in different disciplines.

The core middle grade curriculum can be organized around integrating themes that young people find relevant to their own lives. For example, separate courses in English, arts, history, and social studies might be grouped into the humanities, organized around an integrating theme such as Immigration. Mathematics and science could be combined in the study of themes such as Mapping the Environment.

The core program can capitalize on young adolescents' concerns and curiosity about their own physical and emotional development and their place within the family, peer group, and larger society. Thus the life sciences, which link biology, physiology, genetics, human evolution, and behavior to health education, could be clustered around themes relating to the developing adolescent. Young adolescents are also increasingly concerned about fairness and justice. An interdisciplinary focus on the Struggle for Freedom in America would enable them to explore these subjects in depth.

Interdisciplinary themes can also enhance young people's recognition that the United States is now a part. of a global economy and an interdependent society of nations, and that the composition of our own population is racially and ethnically increasingly diverse. Themes that explore human diversity offer one way of dealing positively with what may become a source of social unrest in this nation as the relative size of minority groups increases in coming years.

## Teaching Students to Learn as well as to Test Successfully

In American education, the assessment of student performance inordinately influences curriculum and methods of instruction. From kindergarten to graduate school, educators teach to the test.

The transformed middle grade school will require new assessment approaches, which are being developed today.[8] In a transformed middle grade school, all forms of student assessment would reflect the purposes of instruction as described in this report. For example, curriculum and teaching methods should promote disciplined inquiry among young adolescents. If higher order or critical thinking is not being assessed, teachers are less likely to demand it of students.[9]

In the transformed middle grade school, tests will more closely resemble real learning tasks. One example is long-term writing assignments that emphasize the development of ideas and coherent expression. A student's written product represents important learning, not memorization, and practice for the test is practice with real writing.[10] Individual and group projects also offer students opportunities to exhibit their knowledge of a subject or issue in a way that is meaningful to them. Problems in mathematics or science that have alternate solutions further represent the kinds of activities that require students to demonstrate a full range of thinking skills, and assimilate ideas and information across subjects.

State and national achievement tests will also differ from current practice. One promising alternative is to assess student performance on the basis of work portfolios as well as of test scores. For example, student achievement in writing and mathematics could be determined partly on the basis of standardized tests measuring state objectives, and partly on examples of students' best work as determined by the youth and their teachers. Such a procedure would establish a direct link between statewide assessment procedures and classroom activities in individual schools.[11]

## NOTES

*Higher order thinking is difficult to define. Resnick has identified the following characteristics: Higher order thinking is nonalgorithmic (the path of action is not fully specified in advance); tends to be complex, involving multiple perspectives; often yields multiple rather than unique solutions; involves dealing with uncertainty and gaps in information; involves the application of multiple, sometimes conflicting criteria; is self-regulating; and involves considerable mental effort.[1]

1. Resnick, L.B. (1987). *Education and learning to think.* Washington, DC: National Academy Press.
2. Keating, D.P. (1988, July). [Cognitive processes in adolescence] (draft). In G.R. Elliott & S.S. Feldman (Eds.), [*Volume on Normal Adolescent Development*] (in preparation). Stanford, CA: Stanford University and Carnegie Corporation of New York.
   Petersen, A.C. (1983). Pubertal change and cognition. In J. Brooks-Gunn & A.C. Petersen (Eds.), *Girls at puberty: Biological and psychosocial perspective.* New York: Plenum.
   Graber, J.A., & Petersen, A.C. (in press). Cognitive changes at adolescence: Biological perspectives. In K. Gibson & A.C. Petersen (Eds.), *The brain and behavioral development: Biosocial dimensions.* New York: Aldine.
3. American Association for the Advancement of Science. (1989). *Science for all Americans: A project 1061 report on literacy goals in science, mathematics, and technology.* Washington, DC: Author.
4. Botvin, G.J., and Wills, T.A. (1985). Personal and social skills training: Cognitive-behavioral approaches to substance abuse prevention. In: C. Bell & R. Battjes (Eds.), *Prevention research: Deterring drug abuse among children and adolescents.* Washington, DC: National Institute on Drug Abuse Research Monograph.
   Flay, B.R. (1985). Psychosocial approaches to smoking prevention: A review of findings. *Health Psychology, 4* (5), 449-488.
   Flay, B.R. (1985). What do we know about the social influences approach to smoking prevention? Review and recommendations. In: C. Bell & R. Battjes (Eds.), *Prevention research: Deterring drug abuse among children and adolescents.* Washington, DC: National Institute on Drug Abuse Research Monograph.
   Glasgow, R.E., and McCaul, K.D. (1985). Life skills training programs for smoking prevention: Critique and directions for future research. In: C. Bell & R. Battjes (Eds.), *Prevention research: Deterring drug abuse among children and adolescents.* Washington, DC: National Institute on Drug Abuse Research Monograph.
5. Botvin, G.J., Baker, E.,Botvin, E.M., Filazzola, A.D., & Millman, R. (1984). Alcohol abuse prevention through the development of personal and social competence: A pilot study. *Journal of Studies on Alcohol, 45,* 550-552.
   Botvin, G.J., Baker, E., Renick, N., Filazzola, A.D., & Botvin, E.M. (1984). A cognitive-behavioral approach to substance abuse prevenetion. *Addictive Behaviors, 9,* 137-147.
   Botvin, G.J., Baker, E., Filazzola, A.D., Botvin, E.M., Danilo, M., & Dusenbury, L. (1985, August). *A cognitive-behaviorial approach to substance abuse prevention: A one year follow-up.* Paper presented at the 93rd annual meeting of the American Psychological Association, Los Angeles.
   McAlister, A.L., Perry, C.L., Killen, J., Slinkard, L.A., & Maccoby, N. (1980). Pilot study of smoking, alcohol and drug abuse prevenetion. *American Journal of Public Health, 70,* 719-721.
6. Schinke, S.P. (1984). Preventing teenage pregnancy. In M. Hensen, R.M. Eisler, and P.M. Miller (Eds.), *Progress in behavior modificiation* (Vol. 16). New York: Academic Press.
7. Resnick, L.B. (1987). The 1987 presidential address: Learning in school and out. *Educational REsearcher, 16* (9), 13-20.
8. Wiggins, G. (1989). Teaching to the (authentic) test. *Educational Leadership, 46* (7), 41-47.
   Archibald, D., & Newmann, F. (1988, August). *Beyond standardized testing: Authentic academic achievement in the secondary school.* Reston, VA: National Association of Secondary School Principals.
9. Keating, D.P. (1988, July). [Cognitive processes in adolescence] (draft). In G.R. Elliott & S.S. Feldman (Eds.), [*Volume on Normal Adolescent Development*] (in preparation). Stanford, CA: Stanford University and Carnegie Corporation of New York.
10. Shepard, L.A. (1989). Why we need better assessments. *Educational Leadership, 46* (7), 4-9.
11. Rothman, R. (1988, October 26). Vermont plans to pioneer with "work portfolios." *Educational Week, 8* (8), 1.

# III-15

# National Science Education Standards: Discussion Summary (Draft)

NATIONAL RESEARCH COUNCIL OF THE NATIONAL COMMITTEE ON SCIENCE EDUCATION STANDARDS AND ASSESSMENTS

Reprinted with permission from the National Research Council, National Committee on Science Education Standards. This draft document is under continuous review and revision.

# National Science Education Standards

## Discussion Summary

DRAFT

National Research Council
*National Committee on Science Education*
*Standards and Assessment*
*(NCSESA)*

September 12, 1994

## What Are National Education Standards?

*National Science Education Standards* offer a coherent vision of what it means to be scientifically literate. The standards describe what all students must understand and be able to do as a result of their cumulative learning experiences. The standards also provide criteria for judgments regarding systems, programs, teaching, and assessment which can provide opportunities for all students to learn science in ways that are aligned with the standards.

The statement, "All students," reflects the social commitment that standards apply to students regardless of background (gender, ethnicity, economic condition), circumstance, or ambition.

"National" means a nationwide agreement, not a federal mandate, on what defines successful science learning and the school practices that support that learning. National standards do not define a national curriculum and do not suggest a form of national standardization of an actual curriculum.

## The *National Science Education Standards* are organized in five categories.

- SCIENCE SYSTEM STANDARDS
  specify the support systems and resources needed to provide all students the opportunity to learn science. They include the criteria that each component and/or level of the education system as well as the whole system must meet and they define how systems function to support the vision of science education presented in the *National Science Education Standards*.

- SCIENCE PROGRAM STANDARDS
  specify the nature, design, and consistency of the school and district science program that affect students learning science.

- SCIENCE TEACHING AND PROFESSIONAL DEVELOPMENT STANDARDS
  specify the criteria for the exemplary practice of science teaching. Such teaching provides students with experiences that enable them to achieve the understanding and ability of the Science Content Standards. Teaching Standards also describe the knowledge and skills that are the foundation of such practice and the professional development of science teachers.

- SCIENCE ASSESSMENT STANDARDS
  specify criteria for assessing and analyzing students' attainments in science and the opportunities to learn which school science programs afford students.

- SCIENCE CONTENT STANDARDS
  specify expectations for the development of: proficiency in conducting inquiry including the use of scientific modes of reasoning, and the ability to apply and to communicate scientific knowledge; scientific understanding of concepts, laws, theories and models; understanding of the interdependent relationship of science and technology; and understanding of the influence of science on societal issues, both contemporary and historical.

## Science education standards are guided by these principles:

- All students, regardless of gender, cultural or ethnic background, physical or learning disabilities, future aspirations, or interest and motivation in science should have the opportunity to attain high levels of scientific literacy.
- The science that all students are expected to learn is defined so that students have sufficient time to develop a deep understanding of essential scientific ideas rather than superficial acquaintance with many isolated facts. All students can attain science knowledge with understanding.
- Learning science is an active process in which all students engage. In science, students ask questions, construct explanations of natural phenomena, test those explanations in many different ways and communicate their ideas to others.
- Science in school will reflect the intellectual tradition -- modes of inquiry, rules of evidence, and ways of formulating questions -- that characterizes the practice of contemporary science.
- Science education reform is a systemic process in which all persons have roles and the responsibility to change.

## *National Science Education Standards Outline*

| | |
|---|---|
| **'Call to Arms'** | |
| **Prologue** | |
| **Table of Contents** | |
| **Reader's Guide** | |
| **Chapter I** | **Introduction** |
| **Chapter II** | **System Standards** |
| **Chapter III** | **Program Standards** |
| **Chapter IV** | **Teaching and Professional Development Standards** |
| **Chapter V** | **Assessment Standards** |
| **Chapter VI** | **Content Standards** |
| **Chapter VII** | **Conclusion** |
| **End Matter** | |

## System Standards

The System Standards are criteria for judging the performance of the parts of the science education system responsible for providing schools with the financial and intellectual resources necessary to achieve the vision contained in the program, teaching, assessment, and content standards.

**System Standards** include:

- **Common Vision.** Policies that influence the practice of science education must be consistent with the program, teaching, assessment, and content standards while allowing for adaptation to local circumstances;
- **Coordination Across the System.** Policies should be coordinated within and across agencies, institutions, and organizations;
- **Continuity.** Policies need to be sustained over sufficient time to provide the continuity necessary to bring about changes required by the Standards;
- **Resources.** Policies must be supported with resources;
- **Equity.** Science education policies must be equitable;
- **Unanticipated Effects.** All policy instruments must be reviewed for possible unintended effects on the classroom practice of science education; and
- **Individual Responsibility.** Responsible individuals should take the opportunity afforded by the standards-based reform movement to achieve the new vision of science education portrayed in the Standards.

## Program Standards

The Program Standards provide criteria for judging the quality of and conditions for school science programs. The program standards focus on issues at the school and district level that relate to opportunities for students to learn and opportunities for teachers to teach science as described in the Standards. The program standards and the system standards address issues related to the vision, support, and integration of the teaching, content, and assessment standards. The program standards address these issues from a point of view close to the experiences of students -- the classroom, school, and district levels. The system standards address them from the point of view farthest from students -- the national, state, community, and the district levels.

**Program Standards** include:

- **Consistency in Program.** All elements of the K-12 science program must be consistent with the standards and with one another and be articulated within and across grade levels to meet a clearly stated set of goals. An effective science program encompasses the elements described below.

- ◊ A set of clear goals and expectations for students is used to guide the design, implementation, and assessment of all elements of the science program.
- ◊ A curriculum framework is used to guide the selection and development of units and courses of study.
- ◊ Teaching practice is consistent with the goals and curriculum framework.
- ◊ Assessment policies and practices are aligned with the goals, student expectations, and curriculum frameworks.
- ◊ Support systems and formal and informal expectations of teachers are aligned with the goals, expectations of students, and curriculum frameworks.
- ◊ Responsibility is clearly defined for determining, supporting, and maintaining all elements of the science program.

- ♦ **Curriculum.** The curriculum in science for all students in grades K-12 should contain the aspects described below.
  - ◊ All the content standards should be included and should be embedded in a variety of curriculum patterns that are developmentally appropriate, interesting, and relevant to students' lives.
  - ◊ Inquiry should be emphasized as a tool for learning science.
  - ◊ The curriculum should connect to other school subjects.

- ♦ **Mathematics and Science.** The science program should be coordinated with the mathematics program to enhance student use and understanding of mathematics in the study of science and to improve student understanding of mathematics overall.

- ♦ **Resources.** The K-12 science program should give students access to appropriate and sufficient resources including time, materials and equipment, space, teachers, and community.
  - ◊ Time is a major resource in a science program.
  - ◊ Conducting scientific inquiry requires that students have easy and frequent opportunities to use a wide range of equipment, materials, supplies, and other resources for experimentation and direct investigation of phenomena.
  - ◊ Collaborative inquiry requires space as well as time.
  - ◊ The most important resource is personnel.
  - ◊ Good science programs require access to the world beyond the classroom.

- ♦ **Equity and Excellence.** All students in the K-12 science program must have equitable access to opportunities to achieve the *National Science Education Standards*.
  - ◊ All students, regardless of sex, cultural or ethnic background, physical or learning disabilities, future aspirations, or interest in science, should have the opportunity to attain high levels of science literacy. By adopting this principle, Standards prescribe the inclusion of all students in challenging science learning opportunities and define a level of understanding that all students should achieve.

  - Not only is science education for all students, but all of the understandings and abilities described in the content standards should be achieved by all students. This is the principle of excellence. The standards describe expectations of achievement.

- **Schools as Communities of Learners.** Schools must be communities that encourage, support and sustain teachers as they implement an effective science program.
  - Schools must demonstrate colleagueship, openness, and trust, as well as explicitly support reform efforts.
  - Regular time must be provided for and teachers encouraged to discuss, reflect, and conduct research around science education reform.
  - Regular time and opportunities within the working day must be provided for teachers of science to engage in professional growth.
  - Teachers need support in creating and participating in networks of reform.
  - An effective leadership structure must be in place that includes teachers.

---

## Teaching and Professional Development Standards

The National Science Standards for Teaching and the Professional Development of Teachers are organized into two sections. The first focuses on what teachers do and is called the "Standards for Science Teaching." The second focuses on how teachers develop professional knowledge and skill and is called "Standards for the Professional Development of Teachers of Science." Together the standards present a broader and deeper view of the science that is fundamental and useful to an informed citizenry; the conviction that scientific inquiry is at the heart of science and science learning; and the need for all students to leave school literate in science.

**Teaching Standards** include:

- **Planning the Science Program.** Teachers of science will plan an inquiry-based science program for their students. In doing this, they:
  - develop a framework of long- and short-term goals for their students,
  - select science content and adapt and design curriculum to meet the particular interests, knowledge, skills and experiences of their students,
  - determine teaching strategies that support the development of student understanding and nurture a community of science learners, and
  - work with colleagues within and across disciplines and levels.

- **Guiding and Facilitating Learning.** Teachers of science will guide and facilitate science learning. In doing this they:
  - interact with their students to focus and support their inquiries,
  - orchestrate discourse among students about scientific ideas,
  - challenge students to take responsibility for their own work and also to work collaboratively,
  - recognize and respond to student diversity and encourage all students to participate fully in science learning, and

    - encourage and model the skills of scientific inquiry as well as the curiosity, openness to new ideas, and skepticism that characterizes science.

- **Assessing Learning and Teaching.** Teachers of science will engage in ongoing assessment of their teaching and of student learning. In doing this, they:
    - systematically gather data on their students and their development,
    - analyze assessment data to guide teaching,
    - guide students in self-assessment, and
    - use student data, observations of teaching, and interactions with colleagues to reflect on and improve practice.

- **Designing and Managing the Physical Environment.** Teachers of science will design and manage a learning environment that provides students with the time, space, and resources needed for learning science. In doing this, they:
    - structure the time available so that students are able to engage in extended investigations,
    - generate a setting for student work that is flexible and supportive of science inquiry,
    - ensure a safe working environment,
    - make the available science tools, materials, print, and technological resources accessible to students,
    - identify and use resources outside the school, and
    - engage students in designing the environment.

- **Building Learning Communities.** Teachers of science will develop communities of science learners that reflect the intellectual rigor of scientific inquiry, the attitudes, and the social values conducive to science learning. In doing this, they:
    - display and demand a respect for and valuing of the ideas, skills, and experiences of all students,
    - give students a significant voice in decisions about the content and context of their work, and require students to take significant responsibility for the learning of all members of the community,
    - nurture collaboration,
    - structure and facilitate ongoing formal and informal discussion based on a shared understanding of rules of scientific discourse, and
    - model and emphasize the skills, attitudes, and values of scientific inquiry.

- **School Planning.** Teachers of science will be active participants in the ongoing planning and development of the school science program. In doing this, they:
    - plan and develop the school science program,
    - have a voice in decisions concerning the allocation of time and other resources to the science program, and
    - plan and implement professional growth and development strategies for themselves and their colleagues.

**Professional Development Standards** include:

- **Learning Science Content.** The professional development of teachers of science requires learning science content through the perspectives and methods of inquiry. Science learning experiences for teachers should:
  - involve teachers in actively investigating scientific phenomena, interpreting their results, and making personal sense of their findings consistent with currently accepted scientific understanding,
  - address issues, events, problems, or topics significant in science and of interest to participants,
  - introduce teachers to scientific literature, media and technological resources that expand their science knowledge and their ability to access further knowledge,
  - build on the teacher's existing science knowledge, skills, and attitudes,
  - incorporate ongoing reflection on the process and outcomes of understanding science through inquiry, and
  - encourage and support teachers to work together.

- **Learning to Teach Science.** Professional development of teachers of science requires the integration of a knowledge of science, learning, pedagogy, and students, and the application of this understanding to science teaching. Learning experiences for teachers of science must:
  - connect and integrate all aspects of science and science education,
  - use actual classroom experience to illustrate and model effective science teaching,
  - address teachers' developmental needs and build on their existing knowledge of science content, teaching, and learning, and
  - use strategies of inquiry, reflection, interpretation of research papers, modeling, and guided practice to build understanding and skill in science teaching.

- **Learning to Learn.** The professional development of teachers of science should enable teachers to build the knowledge, skills, and attitudes needed to engage in life-long learning. Science learning experiences for teachers must:
  - provide regular, frequent opportunities for individual and collegial examination and reflection on classroom and institutional practice,
  - provide opportunities for teachers to receive feedback about their teaching and to understand, analyze, and apply that feedback to improve their practice,
  - provide opportunities for teachers to learn and use various tools and techniques for self and collegial reflection, such as peer coaching, portfolios, and journals,
  - support the sharing of teacher expertise by preparing and using mentors, teacher advisors, coaches, lead teachers, and resource teachers to provide professional development opportunities,
  - provide opportunities to know and have access to existing research and experiential knowledge, and
  - provide opportunities to learn and use the skills of research to generate new knowledge.

- **Developing Professional Development Programs.** Pre-service and in-service professional development programs for teachers of science must be coherent and integrated. Quality programs are characterized by:
  - ◊ clear and shared goals that are based on a vision of science learning, teaching, and teacher development congruent with the *National Science Education Standards*,
  - ◊ integration and coordination of the components so that understanding and skills can be built over time, reinforced continuously, and practiced in a variety of situations,
  - ◊ options that recognize the developmental nature of teacher professional growth and individual and group interests, as well as the needs of teachers who have varying degrees of experience, professional expertise, and proficiency,
  - ◊ collaboration among the people involved in the programs, including teachers, teacher educators, teacher developers, scientists, school administrators, policy makers, and business people-and respect for the unique perspectives and expertise of each,
  - ◊ recognition of the history, culture, and organization of the school environment, and
  - ◊ continuous program assessment that captures the perspectives of all those involved, uses a variety of strategies, both formal and informal, focuses on both the process and impact of the program, and feeds directly into the program improvement and evaluation.

## Assessment Standards

Science Assessment Standards are criteria against which to judge the quality of assessment practices used to measure student attainment and the quality of opportunities provided students to learn science. The assessment standards can be used as guides in developing assessment practices and policy. They are statements of principle identifying essential characteristics of exemplary assessment practices.

**Assessment Standards** include:

- **Coordinating Assessments with Purposes.** Assessments are consistent with the decisions that they are designed to inform.
  - ◊ Assessments are deliberately designed.
  - ◊ Assessments have explicitly stated purposes.
  - ◊ The relationship between decisions and data is clearly stated.
  - ◊ Assessments' procedures are internally consistent.

- **Measuring Student Attainment and Opportunity to Learn.** The achievement data collected focuses on the science content that is most important for students to learn.
  - ◊ Data collected reflect the complexity of the National Science Education Content Standards and measure student achievement of each and all of the dimensions described in those standards.

Opportunity to learn data collected focuses on the most powerful indicators of students' opportunity to learn.

- ◊ Data collected reflects the essential indicators of opportunity to learn as they are described in the National Science Education Teaching, Program, and System Standards.

Equal attention is given to the assessment of opportunity to learn and to the assessment of student attainment.

- ◆ **Matching Technical Quality of Data with Consequences.** The technical quality of data collected is well matched to the consequences of the decisions and actions taken on the basis of the interpretation of those data.
    - ◊ The feature of student attainment or opportunity to learn that is claimed to have been measured was actually measured.
    - ◊ Assessment tasks are authentic.
    - ◊ Time intervals for data collection reflect the time dimension implied in what is being measured.
    - ◊ An individual student's performance is the same on two or more tasks that claim to measure the same aspect of student attainment.
    - ◊ Students have adequate opportunity to demonstrate their achievements.
    - ◊ Tasks and methods provide data that are sufficiently stable to lead to the same decisions if utilized at different times.

- ◆ **Avoiding Bias.** Assessment practices are fair.
    - ◊ Large scale assessments use statistical techniques to identify differential performance among sub-groups of the population assessed that signal potential bias.
    - ◊ Males and females of different racial and ethnic backgrounds have been included in the development of large scale assessments.
    - ◊ Assessment tasks have been reviewed for the use of stereotypes, for assumptions that reflect the perspectives or experiences of a particular group, for language which may be offensive to a particular group and for other features which may distract students from the intended task.
    - ◊ Assessment tasks are modified appropriately to accommodate the needs of students with physical disabilities or limited English proficiency.
    - ◊ Assessment tasks are set in a variety of contexts, engaging to students with different interests and experiences.
    - ◊ Assessment tasks do not assume the perspective or experience of a particular gender, racial, or ethnic group.

- ◆ **Making Sound Inferences.**
    - ◊ When making inferences from an assessment, explicit reference is made to the assumptions on which inferences are based.

## Content Standards

There are eight categories of content standards

**Science as Inquiry**
**Physical Science**
**Life Science**
**Earth and Space Science**
**Science and Technology**
**Science in Personal & Social Perspectives**
**History and Nature of Science**
**Unifying Concepts and Processes**

The first seven categories have standards for grade levels K-4, 5-8, and 9-12. The final category crosses all grade levels.

The content described in the standards does not represent a science curriculum. Curriculum includes not only the content, but also the structure, organization, balance, and presentation of the content. The selection of the fundamental understanding reflected in these standards is based on the following criteria: a) it represents a central scientific idea and organizing principle, b) it has rich explanatory power, c) it guides fruitful investigations, d) it applies to situations and contexts common to everyday experiences, e) it can be linked to meaningful learning experiences, and f) it is developmentally appropriate for students at the grade level specified

------------------------------

**Science as Inquiry** should be recognized as a basic and controlling principle in the ultimate organization of and experiences in students' science education. This standard on inquiry includes both the ability to do inquiry and to understand inquiry.

As a result of activities in grades K-4, all students should develop:

- abilities of scientific inquiry, and
- an understanding about scientific inquiry.

As a result of inquiry-oriented activities in grades 5-8, all students should develop:

- abilities of scientific inquiry, and
- an understanding about scientific inquiry.

As a result of inquiry-oriented activities in grades 9-12, all students should develop:

- abilities of scientific inquiry, and
- an understanding about scientific inquiry.

**Physical, life, and earth and space science** express the primary subject matter of science. Science subject matter focuses on those science concepts, principles, and theories that are fundamental for all students to know and be able to use.

**Physical Science**

As a result of activities in grades K-4, all students should develop an understanding of:

- properties of objects and materials,
- position and motion of objects, and
- heat, light, electricity and magnetism.

As a result of activities in grades 5-8, all students should develop an understanding of:

- properties of matter,
- motions and forces, and
- transformations of energy.

As a result of activities in grades 9-12, all students should develop an understanding of:

- the structure of atoms,
- structure and properties of matter,
- chemical reactions,
- forces and motion,
- conservation of energy,
- the increase in disorder, and
- interactions of energy and matter.

**Life Science**

As a result of their activities in grades K-4, all students should develop an understanding of:

- needs and characteristics of organisms,
- life cycles of organisms, and
- relationship between organisms and environments.

As a result of their activities in grades 5-8, all students should develop an understanding of:

- structure and function in living systems,
- reproduction and heredity,
- regulation and behavior,
- populations and ecosystems, and
- diversity and adaptations of organisms.

As a result of their activities in grades 9-12, all students should develop an understanding of:

- the diversity of organisms,
- the cell,
- heredity,
- nervous system and the behavior of organisms,
- interdependence of organisms in the biosphere,
- matter, energy, and organization of living systems, and
- mechanisms and consequences of biological evolution.

**Earth and Space Science**

As a result of their activities in grades K-4, all students should develop an understanding of

- properties of earth materials, and
- objects in the sky.

As a result of their activities in grades 5-8, all students should develop an understandings of:

- Structure of the Earth System,
- Earth's history, and
- Earth in the Solar System.

As a result of their activities in grades 9-12, all students should develop an understanding of:

- energy in the Earth system,
- geochemical cycles,
- origin and evolution of the earth system, and
- origin and evolution of the Universe.

The **Science and Technology** standard establishes useful connections between the natural world and the designed world and offers essential decision-making abilities.

As a result of activities in grades K-4, all students should develop:

- abilities of technological design, and
- understanding about science and technology.

As a result of activities in grades 5-8, all students should develop:

- abilities of technological design, and
- understanding about science and technology.

As a result of activities in grades 9-12, all students should develop:

- abilities of technological design, and
- understanding about science and technology.

The standard on **Science in Personal & Social Perspectives** connects students with their social and personal world. It helps them develop understandings that enable them to fulfill their obligations as citizens.

As a result of activities in grades K-4, all students should develop an understanding of:

- personal health,
- characteristics and changes in populations,
- types of resources,
- changes in environments, and
- science and technology in local challenges.

As a result of activities in grades 5-8, all students should develop an understanding of:

- personal and environmental health,
- populations, resources, and environments,
- natural hazards,
- risks and benefits, and

- science and technology in society

As a result of activities in grades 9-12, all students should develop an understanding of:

- personal and community health,
- population growth,
- natural resources,
- environmental quality,
- natural and human-induced hazards, and
- science and technology in local, national, and global changes

The standard on the **History of Science** includes an understanding of the nature of science and uses history in school science programs to clarify different aspects of scientific inquiry, science in society, and the human aspects of science.

As a result of activities in grades K-4, all students should develop an understanding of:

- science as a human endeavor.

As a result of activities in grades 5-8, all students should develop an understanding of:

- science as a human endeavor,
- nature of science, and
- history of science.

As a result of activities in grades 9-12, all students should develop an understanding of:

- science as a human endeavor,
- nature of scientific knowledge, and
- historical perspectives.

The **Unifying Concepts and Processes** standard provides students with powerful ideas that help them understand the natural world. These conceptual and procedural schemes are integral to students' science learning experiences. The understanding and abilities associated with this standard should be developed over the entire K-12 continuum.

All students should develop an understanding of and ability to use the following concepts and processes:

- order and organization,
- evidence, models, and explanation
- changes and measurement,
- evolution and equilibrium, and
- form and function.

## *How to Reach Us*

To receive more information on the science education standards, or to be placed on the mailing list for future documents please contact:

**WRITE:**
**National Science Education Standards**
**2101 Constitution Ave, NW**
**HA 486**
**Washington, DC 20418**

**CALL:** (202)-334-1399

**FAX:** (202)-334-3159

**ELECTRONIC MAIL:**
scistnd@nas.bitnet or
scistnd@nas.edu

**We look forward to hearing from you!**

The National Science Education Standards Project is funded by the U.S. Department of Education, the National Science Foundation, the National Aeronautics and Space Administration, the U.S. Department of Energy, the U. S. Department of Agriculture, and the National Institutes of Health.

# III-16

# Why "Outcome" Standards Are Not Enough

LINDA DARLING-HAMMOND
THE NATIONAL CENTER FOR RESTRUCTURING EDUCATION, SCHOOLS, AND TEACHING

The importance of outcome standards and evaluations of school effectiveness has been underscored in recent years. Many policy makers and educators have called attention to the need for better methods of evaluating, analyzing, and reporting student achievements, as well as the importance of encouraging schools to evaluate their practices in terms of student learning. Increasingly, though, the term "accountability" has been associated solely with mandates for student testing and other "outcome-based" standards, without attention to the practices pursued within schools. Some recent proposals have suggested that measures of educational outcomes, such as standardized test scores or school graduation rates, can be used to "hold schools accountable" by triggering rewards, sanctions, or remedial actions. The implication is that accountability is achieved largely by collecting a limited number of performance statistics and using them to "motivate" schools.

One problem with this view is that indicators of student achievement cannot be viewed as purely school outcomes. They are measures of the developed abilities and motivations students bring with them to school, the nonschool learning environments they encounter, and the conditions under which they live, as well as what they learn while they are in school. In addition, aggregated measures of school outcomes fluctuate with changes in school populations and are linked to the supports for children in homes and communities as well as with changes in school practices. Increases or declines in test scores may be due to school variables—changes in school practices, teacher quality, or teaching methods—but they may also be the result of many other factors, including changes in the population of students taking a test.

Genuine accountability—that is, responsible practice and responsiveness to clients—occurs only when a useful set of processes exists for guiding practice and for using information to improve practice (Darling-Hammond and Ascher 1991). Thus, such indicators as schoolwide test scores or dropout rates provide information for an accountability system; they are not the system itself. Outcome standards alone cannot guarantee accountable schools. In fact, if these are improperly designed or unintelligently used, they can actually undermine accountability.

---

Reprinted with permission from *Standards of Practice for Learner-Centered Schools* (New York: National Center for Restructuring Education, Schools, and Teaching, Teachers College, Columbia University, 1992), pp. 7-8.

## Limits of "Outcome" Measures in Assessing School Quality

A key problem with current outcome indicators is that evidence of individual student growth and progress is rarely made available, since student performance data typically are not collected or analyzed longitudinally, either for individuals or for the group of students who have actually remained in a given school or district over a period of years. Common comparisons (for example, between average scores of third grade students in two subsequent years) do not even measure the same children over time, thus they cannot reveal actual progress in student performance, much less suggest whether changes are the result of any shirt in school practice. In contrast to the careful interpretive use of outcome data in other fields such as medicine,[1] the prevailing use of average test score data by schools as a measure of school quality ignores population shifts, fails to measure growth of individual students or cohorts of students, and obscures most important information about either the distribution of student achievement or the effectiveness of the school.

The analysis, use, and reporting of student achievement data must become much more sophisticated, along with the measures themselves. In addition, many other kinds of indicators are necessary for understanding how schools are performing and whether students are benefiting. Performance measures become meaningful in the context of information about students, as well as school processes, activities, and resources. A central reason to evaluate school practices along with student outcomes is that schools' average scores on any measure are sensitive to their student populations, and these populations can shift substantially from year to year. Many schools have high rates of student mobility; in some city schools, pupil turnover exceeds 100 percent annually.[2] Characteristics of student populations can also change dramatically as school neighborhoods change in composition. Some schools serve a constant influx of immigrant students who move on as their families become settled and are replaced by new arrivals to the country. Annual statistical reports for these schools do not represent the progress of a stable group of students. Evaluating the quality of education they provide requires an assessment of the practices teachers engage in, since student achievement within the school cannot be gauged in a stable, comparative fashion over time.

On the other hand, schools of choice, magnet schools, and open enrollment schools often have considerable control over the characteristics of students they will admit. They can influence student outcomes by controlling student admissions. Ostensible improvements in student achievement can be a figment of student selection changes rather than improvements in the quality of education provided. Finally, schools can change their school composition by encouraging or discouraging students from dropping out or transferring to other schools or programs. Schools that successfully persuade struggling students not to drop out can expect their average test scores to decline, while those who push such students out can artificially inflate their schoolwide test scores.

### REFERENCE

Darling-Hammond, L., and C. Archer. (1991). *Creating Accountability in Big City Schools.* New York: ERIC Clearinghouse on Urban Education and NCREST.

---

[1] Patient outcome data are used in much more thoughtful ways in analyses of medical care, for example. The progress of individual patients is evaluated in light of their initial health status and within diagnostic groups over a period of time, with same-patient data providing the basis for longitudinal assessments of treatment appropriateness and outcomes. Judgments about hospital quality use information about standards of care offered, rather than inappropriately aggregated statistics such as average mortality rates (the analog of student average test score reporting for schools), which would penalize hospitals offering care for serious illnesses such as cancer, heart disease and AIDS.

[2] This occurs when there are more student entrances and departures during a single school year—individual students leaving and returning more than once or many students enrolling for only a short period of time before leaving—than the total number of students who are counted as the base student population.

# III-17

## Establishing the Rudder for Change: Creating 21st Century Student Learning Goals

Thomas A. Olson
Northwest Regional Educational Laboratory

Many analysts of effective modern leadership agree with Peter Drucker's assertion that "The best way to predict the future is to create it!" Across the country, communities increasingly believe they can no longer allow their education systems to be merely buffeted by the winds of change swirling around them. There is a renewed belief that we can, indeed, shape education's future so that today's students are equipped with the necessary ingredients for success in the 21st century. And increasing numbers of school systems are beginning to establish processes to consciously create their own futures.

The belief that school systems can shape their own futures through focusing on "21st century" results for students presents a sea of changes in how we think about educational improvement. For decades, the focus in educational planning has been on the opportunities provided by the educational system. Providing a broad and diverse set of curricular offerings, improving textbook selections, ensuring adequate facilities, employing a qualified staff, enriching our library collections—these worthy opportunities for improvement fueled much of our educational work in past decades. State education departments evaluated local systems on their abilities to take advantage of such opportunities. But the energy behind and focus on improvement is now shifting to the results of all these opportunities—success for students.

Manifestations of this major sea of change are seen in the National Goals panel's current efforts to develop national performance standards for students, and a plethora of efforts to set national content standards by groups such as the National Council for Social Studies, the National Council of Teachers of Mathematics, the National Endowment for the Humanities, and the National Academy of Science's Project 2061. States and private accreditation associations are also beginning to shift the focus of their accreditation and school approval procedures to outcomes rather than opportunities. And a small, but increasing, number of local school systems are themselves beginning to establish processes to define expected results for success in the 21st century. All of these state and national efforts, however, will be meaningless unless local systems seriously attempt to create their own rudders to find their way through this sea of change.

## Building the Rudder: An Overview of the Process of Creating the Future

One of the major challenges facing local school systems attempting this major shift in focus is the design of effective processes to help make it happen. Over the past five and a half years the School, Community and Professional Development Program at the Northwest Regional Educational Laboratory (NWREL) has worked with more than 30 local school systems in Northwest states to design, test, and refine such processes. The model for strategic improvement that has emerged is called Creating the Future. (See Blum and Kneidek in the April 1991 issue of *Educational Leadership* for a report of the early pilot test of this model.) Creating the Future rests on the assumption that school systems can, and indeed must, establish agreement on a set of student goals to provide the rudder for restructuring efforts.

The Creating the Future process begins with a conscious agreement by the local board of education and superintendent to restructure the way the shaping of future improvement will take place in the district. This restructuring starts with Peter Block's (1994) notion of "stewardship," which suggests that the real role of modern policymakers and executives is to set up the conditions in which expanding numbers of peoples at all levels can become involved as leaders in shaping the future of an organization.

The board and superintendent initiate the process by appointing a Creating the Future Leadership Team of usually between 10 and 15 people. This group is a "vertical slice" of the district's community and staff, selected to represent a wide range of perspectives (e.g., students, parents, senior citizens, business people, religious leaders, teachers, administrators, and one or two board members). The board also seeks to have the group represent the demographics of the community, so racial and ethnic diversity is a major requirement in forming the team.

While the board desires a wide range of perspectives among the membership of the Leadership Team, it must be careful not to ask any team member to attempt to represent another group. The idea is that the team represents a group of thoughtful people with a rich variety of perspectives, not that the team's membership is an adequate mechanism to represent every perspective in the community. Therefore, the Leadership Team leads an effort to engage staff and community in establishing a set of 21st century goals for students. This set of goals is part of an initial group of policy recommendations presented to the board within the first two to three months of the Creating the Future process. These recommendations, which emerge from the extensive involvement process explained in the rest of this chapter, include (a) values and beliefs about learning and about schools and teaching, (b) a set of 21st century goals for students, and (c) a suggested mission statement for the district (sometimes accompanied by a vision statement).

These "rudders" provide the steering mechanism for the Leadership Team's engagement of staff and community in designing and implementing priorities and action plans for restructuring the system. Throughout this process, the Leadership Team helps ensure the broad vertical slice engagement of others at each stage. In addition, the Leadership Team functions as "keeper of the vision," by helping increase the system's focus on these goals. By the end of 18-24 months, most Creating the Future school districts are well into implementing and monitoring major new strategic priorities to better achieve these student goals. The flow of this work is represented in Figure 1.

In addition to the Leadership Team, the board and superintendent appoint a Research and Evaluation Team. This team, while usually smaller in size, also has vertical slice representation. This group provides assistance to the Leadership Team as it gathers data at each step of the process. The initial focus of the work of the Research and Evaluation Team is to help the Leadership Team assess the effectiveness of an initial round of three broadly representative communications groups meetings to develop the rudder—or guidance system—for the strategic improvement effort.

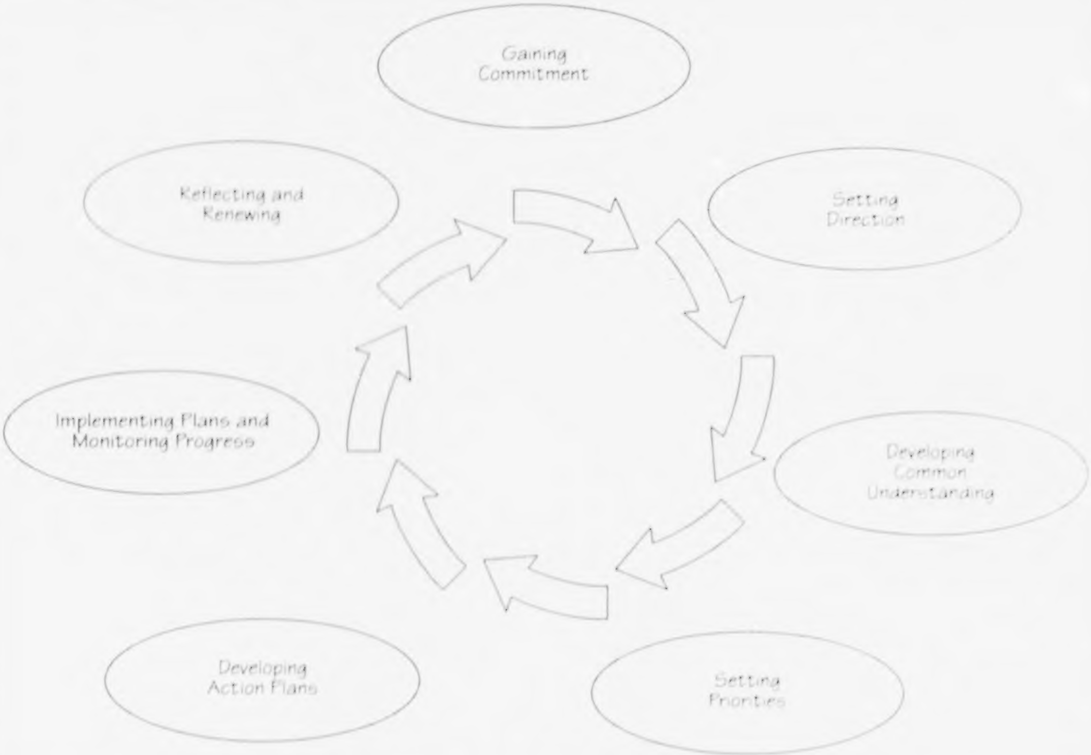
The remainder of this chapter describes how the essential rudder—the set of 21st century student goals—is created. This process models four essential ideas that form the basis for how all succeeding stages of Creating the Future are carried out. These four essential ideas are:

(1) The system needs to focus all efforts on success for students. Reaching greater clarity and agreement on the results wanted for our students is necessary before designing restructured opportunities to achieve the results.

(2) There must be broad engagement and commitment to this focus on students. Broad and diverse participation must occur at every step. This is contrasted with the too-frequent practice of developing plans for restructuring and then attempting to sell them to or get buy-in from others.

(3) When developing recommendations, planning groups, in all cases, should model how an effective learning community works. They should engage in learning from research and others' experiences before developing designs for improvement.

FIGURE 1



This is contrasted with the too-typical mode of planning in which a group convenes and immediately brainstorms solutions from their individual perspectives and biases.

(4) Learning success must be measured and reported; and the quality of measurement of learning success must be improved. Without this focus, the public's demands for greater accountability for results will go unfulfilled.

## Getting Ready to Engage the Community

The Leadership Team's very first work session prepares members for an extensive involvement process to develop the 21st century student goals. Team members first review the basic concepts that frame the overall Creating the Future process. They then engage in some learning and reflections on how people view the future. For example, they review Joel Barker's video, "The Business of Paradigms," and discuss the current paradigms that seem to govern schooling in their district.

For the remainder of the full-day work session, the team develops an action plan to form and work with a large number of communications groups. These groups of 10 to 12 people must also represent a vertical slice of staff and community so the team first considers demographic and other characteristics of community and staff members and derives a set of criteria for the type of diverse representation they hope to see on the groups. Using its membership criteria, the team establishes procedures for inviting people to join the effort, generally agreeing to return for a brief meeting in about two weeks to review nominations for members of the communications groups. The specific number of communications groups established will vary according to the size and complexity of the school district.

Team members then review suggested agendas for three two-hour meetings with the communications groups. They practice, in abbreviated fashion, the various participatory processes that will be used during these three meetings. For example, they practice structured "trio" and large-group discussions of values and beliefs about learning, which will occur in the first communications groups meeting. They also engage in a Jigsaw cooperative learning activity regarding different perspectives on important goals for 21st century success. This activity, which is explained in greater detail later in this chapter, happens during the second set of meetings. The purpose of this practice of structured group process techniques is to equip team members with the skills needed to facilitate the communications groups meetings.

The Leadership Team concludes this first day's work session by developing suggested times and places for the communications groups meetings. They also develop specific facilitation assignments, usually "pairing up" two members to facilitate each communications group.

At this point, the Leadership Team has developed the basics of a plan to engage anywhere between 100 to 500 community and staff members in generating ideas for necessary policy directions for the future (values, student goals, and a district mission and vision).

## Many Hands to Build the Rudder: Communications Groups Meetings

### Meeting #1: Orientation and Identifying Core Values

The first two-hour meeting of communications groups (often with all groups coming together in the same place on the same evening) provides an overview of the basic ideas and work schedules for Creating the Future. The early part of the meeting also establishes a common ground of understanding about the need to increase the focus on 21st century success for all students. For example, many leadership teams show a videotape highlighting the rapid changes being experienced in the country's social, technological, and economic fabric.

Each communications group engages in a structured small-group/large-group process in which they identify core values and beliefs about (a) students and learning and (b) schools and teaching. Common areas of agreement are summarized by the Leadership Team facilitators. When closing the session, the Leadership Team members agree to prepare a report about the session before the second two-hour meeting.

Evaluations from Creating the Future districts reveal that as people leave this first meeting they feel energized and appreciative of the opportunity to participate in shaping future directions. Illustrative comments from evaluations include:

> "They're really listening."
>
> "This is the first time I've been given the opportunity to discuss core values and beliefs about learning with school officials."

And, from a student member of a Leadership Team:

> "I didn't really believe all those adults would let me facilitate the discussion; but they did! And they respected my leadership!"

Indeed, in many cases, people have stayed long after adjournment in parking lot discussions to review the ideas and themes generated during the meeting.

### Meeting #2: Identifying 21st Century Goals for All Students

The second meeting, also about two hours in length, gives group members an opportunity to identify the goals they believe are necessary to help students achieve success in the 21st century. Districts generally report success with holding this meeting shortly after the first meeting, often no more than one week later. This helps sustain the momentum, enthusiasm, and attendance.

After reviewing the written report from the first meeting, participants immediately engage in a cooperative learning activity to help them study a variety of perspectives about important 21st century goals for student success. This activity, called a Jigsaw, asks people to organize into home groups of four or five people. A series of articles is distributed in each group. Each group member selects a different article to read and becomes the home group's expert on that article. That expert will be the only one in the group who has read the article, and will be expected to give a 3- to 5-minute report to the group on the essential ideas it contains. Before they do this reporting, the experts on each article gather together. In this way, experts from different groups who have read the same article can advise each other on what basic ideas ought to be in the reports back to the home groups.

The home group reporting then takes place. Leadership Team members facilitate this process and act as time keepers to ensure that all articles are reported in each home group. Then, after the reporting, the home groups discuss and record on flip charts their perspectives on important 21st century outcomes. These then are reported to the total communications group. Leadership Team members summarize and test whether there is agreement on common themes. They then conclude the session by promising to bring a report of this work session back to the third meeting (often scheduled for the very next week). After quickly reviewing the nature of the agenda for the third meeting, the facilitators adjourn this meeting. Again, participants in virtually every district have reported a high level of excitement and enthusiasm about the ideas shared in this session and a renewed belief that the district really cares about their opinions. The "learn first, then decide" Jigsaw activity stimulates thinking, and a number of participants have reported that it helps broaden their perspectives about necessary goals for success.

It is very important to select a good balance of articles for the Jigsaw activity. For example, one or two articles might focus on the "world of work" perspective. Possibilities include the summary of the SCANS (Secretary's Commission on Achieving Necessary Skills 1992) report from the Secretary of Labor or an abstract of *Workplace Basics: The Skills Employers Want* (Carnevale, Gainer, and Meltzer 1988; see excerpt in this section), published by the U.S. Department of Labor. But, it is important to balance this perspective with others. For example, Ernest Boyer's article "Civic Education of Responsible Citizens" in this section is an excellent counterpoint to the world of work perspective. His notion of "civic duty" provides group members with a powerful concept as they think about what goals are important for students to achieve.

Other articles useful in this Jigsaw activity come from the North Central Regional Educational Laboratory's *Guidebook-1: The New Definition of Learning: The First Step for School Reform* (see Jones and Fennimore 1990), and Alfie Kohn's (1991) "Caring Kids: The Role of Schools" (in this section), which adds a perspective on caring for and respecting others. Articles from the more popular periodicals also prove helpful. For example, Kenneth Labich's "What Our Kids Must Learn" from the January 27, 1992, issue of *Fortune* magazine presents, in lay terms, a broad array of necessary attributes for success in an increasingly interdependent and technological future. None of these articles is sufficient by itself to describe the full set of goals necessary for success, but they do help each communications group consider a wide array of perspectives before they advise on the goals.

Often, after the second meeting, the Leadership Team will convene to prepare a synthesis of the input from all of the groups. What they discover is generally a natural grouping of 8 to 10 major goal statements. They often bring this synthesis (along with each group's report) back to the third meeting.

### Meeting #3: Considering Our District's Mission

Having provided input for a statement of core values and beliefs and a preliminary set of student learning goals, the communications groups now use these statements as a foundation for thinking about the very purpose of the school district. In essence, they are asked to consider the question, "If we truly believe in these values and strive for these results, what should be the business of the school district?" A well-developed mission statement provides a basic, brief, written statement that clearly communicates the answer to this question.

The following basic question stimulates the structured group interaction: "Given our values and desired student goals, should the purpose of our district be to (a) provide a rich array of learning opportunities or (b) to ensure learning success for our students?" This dichotomy, which captures the essence of a focus on opportunities versus a focus on student goals stimulates healthy discussion in communications groups. Each group weighs the relative advantages and disadvantages of each type of mission statement and reports the results of their discussion and whether they were able to achieve consensus on which mission is the most appropriate one for their district.

There is obviously no wrong or right answer to this discussion question, and sometimes consensus is not achieved. Some leadership teams have, appropriately, simply reported this lack of consensus to the board and suggested that this is an issue that the board must decide on. Of course, they always include the nature of their discussion and their weighing of the pros and cons of each of the two alternatives.

An optional final activity in this third work session is the construction of a vision statement. For example, in the North Slope Borough School District, headquartered at the "top of the world" in Barrow, Alaska, one member of the Leadership Team, Cora Sakeagak, considered the reports of the discussions held in three meetings in each of the eight villages across the district's 88,000 square miles of tundra. Much of the discussions about values, goals, and mission had centered on the growing consensus among villagers and staff that students needed to be prepared to succeed in both "worlds," the primarily subsistence economy of the villages and the world outside the district. She spent a Friday evening thinking about this and offered this statement to the Leadership Team at its Saturday morning session:

> Here is a young man graduating from Trapper School. He graduates with marketable skills; indeed, he has been accepted at two top colleges in the nation!
>
> This child throughout his career in the Trapper School system has demonstrated the ability to tell right from wrong in the classroom as well as in the community. He is an active member of the church youth group, acting as a big brother and peer counselor to younger students.
>
> His love for learning has earned him the Alaska Decathlon silver medal, as well as "Alaskan of the Year" Scholarship. With his skills in tact and the ability to mediate he single-handedly prevented a major split between the association of student government and AFN Youth Council.

BEST COPY AVAILABLE

> A lifelong member of his father's whaling crew, he has learned his role in the crew vital to a successful harvest. As well, he demonstrates his deep commitment to his heritage as a member of the Nuiqsut dance group and the ability to communicate in Inupiaq.
>
> When asked what was the most important thought to himself is, he said, "I think I am important to myself and my family and my people. I am Inupiaq. I know my responsibility to my people and the family I hope to have one day. Thank you to my whole family—parents, grandparents, aunts, uncles, the school system for opening every opportunity to me. My goal now is to complete college and come home to be a productive member of Nuiqsut.

The Leadership Team said this statement captured much of what they envisioned. The district now uses this statement (often translated in Inupiaq) as a way to convey to many people the hope and promise of focusing on student success in the 21st century. Such vision statements help make real the essence of the Creating the Future process and the consensus agreements that flow out of it.

### After the Three Meetings: Now the Hard Work Begins!

Soon after the third round of communication group meetings, the Leadership Team convenes to develop a final draft of a recommended values statement, student goals, and mission statement for adoption by the board of education. The team often asks communications groups to review the draft one more time before carrying it to the board for approval. Once the board adopts the statements, there is a tendency for the team to breathe a large sigh of relief and believe that their work is done. But, of course, the work has only just begun.

The next task is to ensure that there is expanded "common meaning" about the 21st century goals among the staff and community. To begin work on developing and expanding this meaning, the Leadership Team now must lead a process (again, with vertical-slice representation) to develop written, narrative descriptions of the qualities and characteristics of people who are achieving the 21st century goals. In our Alaskan sites, for example, this process began with village meetings involving elders, parents, students, and staff who discussed the characteristics of members in the village (adults or students) who are good examples of people achieving these goals. In other places, broadly representative study groups have convened to develop such descriptions. These common descriptions provide the basis for expanding common meaning across the system. Equally important, they provide the basis for the even harder work to which this book is devoted—improving the assessment of the goals for 21st century success.

## What Have We Learned About Goal Setting Through Creating the Future?

With every district in which we work, we learn more about the power of broad involvement in goal setting. The following are some of the essential learnings we have realized so far:

(1) People are anxious and eager to participate in such a process. Nevertheless, special effort by the Leadership Team to make personal invitations is absolutely critical to ensure a high participation rate in goal-setting meetings.

(2) Such meetings as described here seem to maintain the best participation rate when held in a relatively compressed time period (e.g., three meetings held over one month).

(3) The meetings will be successful and productive only if the structured group process is followed and if each meeting is held to about two hours. Equally important is documenting the results of each meeting and sharing these results at the next meeting. This allows participants to see themselves and their ideas in the minutes of each meeting.

(4) Special logistical attention helps ensure success. Plenty of advance notice of the meetings, name tags, participant warm-up activities, inexpensive food or snacks, and a nice place to meet are all secrets to successful goal-setting meetings. Together, they underscore the importance the district attaches to this work.

(5) People can and do reach consensus on important core values about learning and goals for 21st century success! This realization by participants is a major energizer.

(6) Some groups' members, with help from national organizations, are increasingly resisting attempts to get greater clarity about goals for our students. (See the March 1994 issue of *Educational Leadership*.) These groups criticize efforts such as the one described here as "fuzzy-headed, outcome-based education." They are concerned that the schools are inappropriately taking over parental responsibilities and are using this kind of goal setting as a smoke screen. Others charge that the core belief that all students can succeed will only "dumb down" education to the lowest common denominator. Their arguments are often emotionally charged and their tactics can sometimes be disruptive to a consensus-building process.

Many of the districts with which we have worked have experienced such challenges. Those that have successfully moved ahead did so by encouraging dialogue and openness with these skeptics throughout the goal-setting process. At the same time, they were very well prepared to keep the process on track, and not let a small vocal minority derail it. Those that welcomed the dialogue—within the structured process—found that it enriched everyone's understanding. Those that attempted to change the process totally to accommodate this small minority often experienced an erosion of participation by the majority, and, in at least one case, a cessation of the goal-setting effort. In coping with this issue, each leadership team's skill at being prepared, good listeners while at the same time remaining firm about staying with the agenda was the key to success.

(7) It is very difficult for a school system to reorient itself to continuous vertical-slice participation in planning strategic improvements, but broad involvement in goal setting is highly successful. Boards like it because it yields a large and enthusiastic group of people who feel ownership. Yet, after the goals are adopted and new forces intervene (for example, new state mandates, pressure group demands, collective bargaining agreements), there is a tendency for systems to fall back into the traditional mode of limiting involvement in planning improvements to relatively small groups of administrators, board members, and teachers and attempting to sell such plans to the rest of the staff and community. The Leadership Team should be a mechanism for continuing to build the system's capacity to operate in a more open, inclusive fashion. After being involved in what we can safely predict will be a successful experience in leading broad involvement in goal setting through the Creating the Future model, the Leadership Team can train others in how to do business in new and different ways—ways that ensure broad involvement and commitment to success for all students as they navigate into a future of unprecedented change.

## REFERENCES

Block, P. (February 1994). "Reassigning Responsibility." *Sky*, 26-29.

Blum, R., and A.W. Kneidek. (April 1991). "Strategic Improvement that Focuses on Student Achievement." *Educational Leadership* 48, 7: 17-21.

Boyer, E.L. (November 1990). "Civic Education for Responsible Citizens." *Educational Leadership* 48, 3: 4-7.

Carnevale, A.P., L.J. Gainer, and A.S. Meltzer. (1988). *Workplace Basics: The Skills Employers Want*. Alexandria, Va.: American Society for Training and Development and Washington, D.C.: U.S. Bureau of Labor, Employment and Training Administration.

Kohn, A. (March 1991). "Caring Kids: The Role of the Schools." *Phi Delta Kappan* 72, 7: 496-506.

Jones, B.F., and T.Fennimore. 1990)."What is the New Definition of Learning?" In *Restructuring to Promote Learning in America's Schools: A Guidebook-1: The New Definition of Learning: The First Step for School Reform*. Oak Brook, Ill.: North Central Regional Educational Laboratory.

Labick, K. (January 27, 1992). "What Out Kids Must Learn." *Fortune*.

Secretary's Commission on Achieving Necessary Skills. (April 1992). *Learning a Living: A Blueprint for High Performance: Executive Summary. A SCANS Report for America 2000*. Washington, D.C.: U.S. Department of Labor.

# III-18

# Developing Exit Outcomes by Community Summit

GARY E. DUNKLEBERGER
CARROLL COUNTY PUBLIC SCHOOLS, MARYLAND

In the Carroll County Public Schools, clear, broad-based agreements have been reached regarding what students should learn through the schooling process. District leaders brought together various stakeholders to develop "exit outcomes"—statements of what students should learn before they leave high school. These outcomes are becoming the focus of all curricular, instructional, and assessment activities.

The degree of success a school system can achieve in using exit outcomes to transform the educational process depends on the extent to which the outcomes are credible to and owned by various stakeholders. This issue of ownership is crucial; it often makes the difference between acceptance and rejection of outcomes. Therefore, outcomes for a school system must belong to the entire community—not merely the system.

The key issue, then, is how to build a broad-based consensus that involves all stakeholders in the community. Carroll County Public School District leaders decided to conduct an "Exit Outcomes Summit" in which large segments of various constituent groups were assembled to forge a set of outcomes for the system. A committee of education professionals, parents, and business representatives was charged with oversight of this process. Invitations to participate in the Summit were sent by the superintendent to all teachers and administrators in the system; their involvement was regarded as crucial to success, and district leaders wanted to dispel any notion teachers might have that establishing student outcomes was something being "done to" them. Officers from the parent-teacher organizations in each of the 33 schools were also invited, as were members of the business community, political leaders, county Curriculum Council members, representatives from each bargaining unit, board of education members, and student government representatives. Over 2,000 invitations were extended and 723 individuals indicated their willingness to participate.

The challenge was to find a way to provide meaningful involvement for so many participants. We were concerned that the process might degenerate into chaos. Clearly, a group this large could not function as a committee to hammer out what students should be able to demonstrate prior to finishing their schooling experience. We decided to segment the task into three phases, which would be accomplished in five successive half-day sessions during one month.

During the first phase, participants generated draft outcomes. In the second phase, descriptors were added to expand and clarify the meaning of each outcome. These two phases were conceived as building stages; any idea that received consensus support in one or more of the approximately 20 working groups was added to the draft document. While this served to capture all of the ideas generated, it failed to distinguish between those things that were the beliefs of a small segment of individuals, and those that might have had more universal support. As a result, in the third phase we attempted to find out what support existed for each of the draft ideas. Specifically, we looked for those things for which there was 80 percent agreement among the participants.

The committee that had been established to develop these exit outcomes had done some work in advance of the Summit. Based upon the SCANS Report (1991), *Workplace Basics: The Skills Employers Want* (Carnevale, Gainer, and Meltzer 1988), and a review of the futurism literature, a preliminary list of exit outcomes was developed for use in the Summit.

## Project Phases

### Phase I

Prior to Phase I, all participants in the phase were mailed a packet of information. This packet, containing the resources cited earlier, was to be reviewed prior to participation so that Summit participants could begin their work with a common background. Participants also received an overview of the process being used in Carroll County, sample exit outcomes from three other systems, and some background information on how an outcomes-driven educational system might work.

To provide focus for their review of the materials, participants were asked to synthesize the information and their own thinking into two short lists: one describing the most important changes in the 21st century and another identifying the most significant ways education will need to change in response.

So that all participants would have an appropriate focus on what students should learn to be prepared for the 21st century, each session in Phase I began with a three-part overview. First, facilitators reviewed the direction in which the school system was headed, then participants held a discussion of outcomes based on the materials in their packets and reviewed futuristic conditions that will influence the skills needed for today's students to enter tomorrow's world of work. That done, participants began the process of developing potential exit outcomes for Carroll County students. Groups of six or eight were asked to brainstorm things students should be able to know and do and attributes they should have before concluding their education. They were told not to worry about phraseology, just to capture the concepts in a word or two. Because the ideas they were developing were almost like trial balloons, each group was asked to record the things upon which members were able to agree on paper cut-out balloons. The small-group discussions that ensued were lively and sometimes heated. Because groups were to write their ideas large enough to be read, the size of the paper balloon cut-outs virtually eliminated any debate on wording and focused attention upon the concept involved. In the end, however, each of the working groups generated anywhere from 4 to 10 ideas for potential outcomes.

A conscious decision was made to share the committee's draft outcomes with the group only after they had generated their own ideas. We believed that to do so earlier would only focus attention upon the preliminary draft and engage the groups in a discussion of whether or not they agreed with it. Because the purpose of this phase was to generate and capture all possible outcomes, those drafted by the committee were displayed on three-foot, laminated paper balloons once the groups had finished their work.

Groups were then asked to match the ideas they generated with the draft. If any of a group's ideas matched one of the seven draft outcomes from the committee, the group was asked to tape its balloon on the larger balloon for that outcome. If there was some overlapping but not a complete match of a group's idea and a draft outcome, the group was asked to tape its smaller balloon on the edge of the larger one. If an idea did not match with any of the draft outcomes, the group was asked to tape its balloon and idea up on a separate place on the wall.

Working groups engaged in animated discussions as to whether the committee's draft outcome statements captured what they had intended in their own ideas. The room was filled with activity. Some groups reached agreement on what to do with all their balloons before leaving the table to tape them on the larger balloons. Other groups placed each on the wall as they reached a decision on how to handle it. Amazingly, when all the groups were finished, nearly all their outcome ideas were attached to larger balloons. Because their work so closely matched that done by the much smaller committee, we were able to confirm that there was substantial consensus on those seven outcomes.

A number of balloons from the working groups, however, did end up in the place designated for those which did not match any of the draft outcomes. The committee then went about the task of incorporating each of these ideas into the draft document. A number were infused as descriptors of various outcome statements. Additionally, two new outcome statements were developed. One of these related to basic skills, the other to the idea that every student should know a second language.

This process allowed for the productive involvement of over 200 individuals in the development of exit outcomes for the district's students. Just as important, it allowed participants to generate and systematically sort ideas in a way that led to a clearly defined product. Evaluations completed by participants indicate that, in general, they appreciated and enjoyed the opportunity to be involved.

### Phase II

The second phase of this Summit was designed to develop indicators that would add "meat" to the outcomes. This was considered crucial because these indicators give greater specificity and meaning to broad outcomes such as "Perceptive Problem Solver" or "Lifelong Learner." These indicators became part of the foundation for the framework by which performance can be assessed.

Over 200 people agreed to participate in two separate sessions in Phase II. As with the first phase, participants were mailed a packet of materials for review in advance, each session began with an overview of materials and concepts, and considerable pains were taken to ensure that each group of six or eight represented a balanced cross-section of the various stakeholder groups participating in the summit process. Participants were assigned to specific groups to eliminate the likelihood of people working solely with individuals they already knew. It was considered important to ensure, for instance, that no group consisted only of teachers and lack the valuable parent and business community perspectives.

Each group of six or eight participants was charged with reaching consensus on the indicators for two outcomes developed in Phase I. As a group, they were able to add, delete, or modify items. Once consensus was reached, each group was paired with another that had worked on the same two outcomes. They were asked to develop a composite consensus within the combined group.

For most of the groups, this was a challenging activity. Each of the combined groups now contained 15 or 16 people, which made the group dynamics more difficult. Even so, the process worked well. Having had the opportunity to express their views and contribute their ideas in the smaller groups made the combined group setting more manageable. In all cases except one, the combined groups were able to reach agreement on a joint report. In the one exception, the two initial groups simply had such diametrically opposing views that they could not come to an agreement and issued two separate minority reports.

Use of the process resulted in deletion of two outcomes developed in Phase I by some particularly strong-willed individuals—outcomes related to learning of the basics and the knowledge of a second language by all students. It was generally viewed that the basic skills were already included in the outcome regarding development of "Able Communicators" who can convey information in the written, spoken, and numerical form. Some participants in the first round, however, felt that a basic skills outcome needed to be set aside by itself. The outcome related to second language learning for all students was presumably championed by foreign language teachers who were a part of Phase I of the Summit. This caused much agonizing by steering committee members who did not philosophically support inclusion of a second language outcome but who felt it important to remain true to the process of including everything agreed upon during the Summit—without regard to the judgments of the committee or the system leadership.

It is interesting to note how the Phase II groups assigned these two particular outcomes dealt with them. At the first day's session, two groups were each given one of these outcomes and another two groups were given the other. Each was charged with developing indicators for their particular outcome. Three of the groups simply refused to undertake the task. The fourth completed the task but registered its objection. When the two groups that had the basic skills outcome were then combined, they quickly agreed that it should not be a separate outcome and that it was already incorporated in "Able Communicators." They felt that basic knowledge was a prerequisite to being able to communicate information.

The groups that had the second language draft outcome quickly agreed that it was not a legitimate expectation. Group members felt that for many students—especially those who are college bound—this is a realistic goal; however, it was judged not to be a proficiency that all students should have to demonstrate before completing their schooling experience.

The following day, the second session was conducted with an entirely different group of individuals. Virtually the same scenario unfolded with these two outcomes; four different groups found each of them to be unnecessary. As a result, both of these outcomes were dropped from further consideration. The summit process had screened out material that had been introduced by a minority of individuals or special interest group.

### Phase III

Phase III of this summit process was intended to identify those draft outcomes for which there was widespread agreement. Specifically, this phase was designed to determine which statements were supported by at least 80 percent of Summit participants.

Each of the 27 facilitators led a group of six to eight participants through a discussion of each outcome and indicator, attempting to focus upon strengths and potential problems. At the conclusion of the discussion on each outcome, the facilitator asked members to vote as to whether or not they agreed with the outcome and indicator. Although participants had two-and-a-quarter hours for this task, it was evident that most groups would not be able to deal with all seven outcomes. To avoid the risk of getting lots of data on the first outcomes and very little on the rest, groups were asked to begin with different outcomes. As a result, although most groups were able to discuss only four or five of the outcomes, adequate data were obtained for all.

All outcomes received at least the 80 percent level of agreement established as the criterion. In fact, all were supported by more than 90 percent of participants, except for the outcome related to "Individuals with a Positive Self-Concept," for which there was 83 percent agreement. Much of the debate on that outcome related to whether or not the school should be involved in developing self-concept and whether or not various aspects of it were measurable. Five of the remaining 6 outcomes were supported by at least 97 percent of those in attendance.

About a dozen of the indicators for various outcomes received less than the required 80 percent support needed and were removed from the document. Most of these had to do with the outcomes of "Individuals with a Positive Self-Concept."

## Evaluation

We were very interested in participants' feelings and impressions after the sessions. We believed that the entire summit process would have been a counterproductive endeavor if individuals left with a high frustration level and felt that they had not played an important role in developing the district outcomes.Therefore, participants were asked to provide feedback at the end of each session with regard to seven questions using a four-point Likert scale. In addition, one open-ended question invited participants to share what they thought were the strengths and weaknesses of the summit process. In all five sessions of the three phases, responses indicated that participants viewed the sessions as well organized and successful in meeting their objectives. More important, the most highly rated item was the one asking participants to indicate their happiness about having had the opportunity to participate in the Summit. Individuals were very positive about being able to contribute and voice their opinions.

## Follow-Up

An important aspect of this summit process was the way participants were provided summaries of the conclusions that were reached. At the end of each round, an analysis of the issues and final consensus decisions were mailed to all participants, along with the specific changes made in the draft text. This allowed those involved to clearly see what they had accomplished, and was particularly important because there were two different sessions in Phases I and II. This analysis was also mailed to all participants in previous sessions so that they could follow the changes as they were being made.

Participants indicated that this form of follow-up served a number of important functions. While providing a summary of what had been done, it also allowed individuals to see that their ideas were given respectful treatment and revealed exactly how and why some ideas were eliminated. The summaries also conveyed a very clear message about the importance of the task.

The summit process is an effective means for developing exit outcomes. By involving large segments of various stakeholder groups, the process served to achieve consensus—within the educational community and the community at large—about what should be expected of our schools. At no time before had the community debated educational issues as vigorously as they did during the month of the Summit. This was clearly reflected by the great number of articles carried in the local press—often on the front page with banner headlines. The outcomes that emerged from the Summit truly express the will of the people of Carroll County.

## REFERENCES

Carnevale, A.P., L.J. Gainer, and A.S. Meltzer. (1988). *Workplace Basics: The Skills Employers Want.* Alexandria, Va.: American Society for Training and Development; Washington, D.C.: U.S. Department of Labor, Employment and Training Administration.

Secretary's Commission on Achieving Necessary Skills. (1991). *What Work Requires of Schools: A SCANS Report for America 2000.* Washington, D.C.: U.S. Department of Labor.

# III-19

## Outcomes, Performances, and Assessment: The Agony and Ecstasy of Development

TERRY MAZANY
OAKLAND UNIFIED SCHOOL DISTRICT

PAMALA NOLI
CALIFORNIA SCHOOL LEADERSHIP ACADEMY, CALIFORNIA DEPARTMENT OF EDUCATION

Have you ever read a journal article describing the remarkable transformation of a school or district and marveled at the wisdom and competence the participants must have possessed? The descriptions often create the image of a sleek, well-managed ship sailing flawlessly on course from port to port. Do you then compare your own attempts at change and wonder why your metaphor is more like a rocky auto trip on undeveloped back roads full of detours and roadblocks? Have you wondered if you and your colleagues constitute a small minority who lack the magic knowledge that allows others to take such picture-perfect trips?

Over the past three years we have worked with a variety of schools engaged in redesigning teaching, learning, and assessment in outcome-driven ways. Without exception, these schools were led by teams of exceptionally talented, technically skilled, caring, and committed professionals. And, yet, their journeys were all more closely reflective of the rocky road metaphor than that of clear sailing. The members of these school communities can be characterized as explorers struggling to chart a meaningful course. Their struggles to make personal meaning out of such notions as student outcomes, performance-based learning, performance indicators, standards, rubrics, and portfolios, to name a few, are slow, frustrating, and full of detours.

Our multiple experiences led us to conclude that significant change efforts are about wrong turns, dead ends, and backtracking. In those situations where the travelers do not succumb to the urge to turn back, fearing that they aren't up to the challenge, the journey can become transformative and synergistic. These explorers' travel logs and our observations have led to some process lessons that provide a useful compass for pioneering schools whose members are charting a course for school reform.

## Lessons Learned

Of the many lessons we have learned, five are particularly instructive.

First, developing outcomes is an exercise in creating meaning. The community of learners is creating new and shared insight into what it means to be successful and what quality means. Accomplishing change of this magnitude calls for processes of inclusion in decision making built on a foundation of growing trust and an understanding that anxiety and frustration are natural to the process.

Second, a school cannot become outcomes driven without the concurrent redesign of assessment and curriculum. This reality reflects the goal that assessment and learning become seamless. Outcomes define performances. Performances define assessments. Assessments inform outcomes, which redefine curriculum and performances. This lack of linearity almost ensures back-and-forth swings between the development of outcomes, the demonstration of performances, and experimentation with assessments. All this commotion is accompanied by predictable discomfort for all except the most "abstract-randoms" among us.

Third, the definition of outcomes is a process of continuous improvement. Learner outcomes and their measurement continue to be refined as schools better understand and determine their expectations and experiences with the measurement of outcomes.

Fourth, the development of outcomes and authentic assessment has profound implications that affect the values and norms of a school culture. Explicit in the practice of authentic assessment is a defining commitment to honesty and integrity, which creates a dissonance with inauthentic practice and behavior. This dissonance sends ripples across the cultural landscape of the school, enhancing the integrity of that community. For instance, schools with defined outcomes can inform students up front of achievement expectations, eliminating the "gotcha" strategy of "answer-hiding" testing.

Fifth, and finally, if outcomes are defined in the context of a community of learners, then the outcomes should be relevant to all members of the community and not limited in their application strictly to students. All members of a community of learners must be accountable for their own personal achievement of the outcomes!

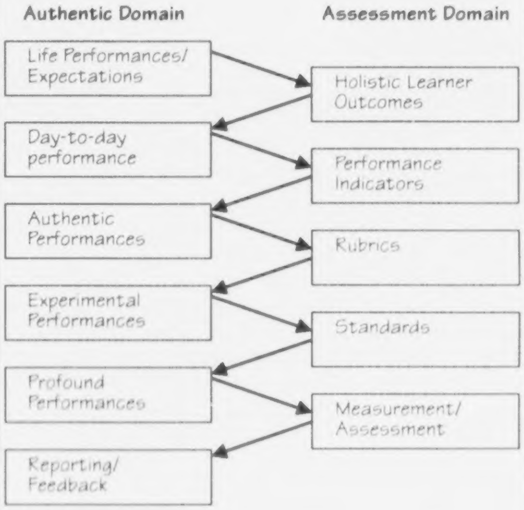
## A Process Model for the Development of Outcomes

The journeys of several schools provide insight into the process through which schools define outcomes and use those outcomes to guide the redesign of teaching, learning, and assessment. The particular journey of Starlight Elementary School in Watsonville, California, provides a concrete example of the discovery that the change process is not a linear sequence progressing from creating outcomes, to rewriting curriculum, to designing assessment, and then to designing performances of student work. Starlight School, along with other inquiring schools, learned that each construction informs and potentially changes previous understandings. The determination of outcomes can become a remarkable catalyst for the construct of authentic student work as performance assessment becomes the curriculum. Figure 1 models the developmental experiences of Starlight School and several others.

FIGURE 1

A PROCESS MODEL FOR THE DEVELOPMENT OF OUTCOMES

*The experiences of Starlight can be modeled with the following diagram:*



The point of entry into this process, or where a school begins, appears to have little effect on the school's ability to succeed. Some schools have begun with an attempt to implement student portfolios (someone said it was the trendy thing to do) and soon realized that without defined student performances based on outcomes, portfolios were little more than tidy collections of stuff. Others entered through the door of authentic assessment and, after much struggling, concluded that without the construction of desired performances based on valued outcomes, authentic assessment led only to a fragmented series of performances disconnected from the curriculum. Still others began by designing integrated curriculums or thematic units that generated terrific performances, but about what? After all, the products of thematic units are not deemed relevant according to the yardstick of a standardized test.

Starlight School entered the process guided by a simple idea: Before we can talk about questions of how to get there, we must first be clear about where we want to go and where we are right now. For the members of the Starlight School community, the starting point was changing their conception of outcomes.

*Lesson Learned: Outcomes act as a "north star," guiding the work of school charge. This journey can start at any point on the map, but until you take out your compass or look up at the stars, you will wander aimlessly in the wilderness reacting to wave after wave of school reform fads.*

### Life Performances and Expectations

Although Starlight School's entry point was the definition of holistic student outcomes, school members experienced great frustration with their construction. Laborious hours of synthesizing brainstormed lists seemed to heighten school members' discomfort with what they were producing. One teacher expressed her anxiety with being responsible for outcomes based solely on a self-generated list by asking the faculty members who had made them omnipotent. At that point, faculty members asked themselves how they might justify their brainstormed outcomes. This question shifted their focus outside the school to those life performances and expectations their students would someday face.

Life performances are those public and private acts that constitute success in the next school experience and later life experiences. Life performances can range from interpersonal communications needed for a healthy relationship to a knowledge of systems required for the workplace. Authentic life performances demonstrate a display of understanding and the creation of new knowledge in new contexts. Life expectations also include those "tests" we meet in our childhood and adult lives, ranging from job interviews and driver's license tests to merit badges, CPA certifications, and Bar examinations.

Starlight School began with a study of *Learning a Living: A Blueprint for High Performance,* the report from the U.S. Department of Labor's Special Commission on Achieving Necessary Skills (1992), which provides a basis for defining the competencies and skills required for the world of work by identifying a three-part foundation and five competencies required for our workforce. Starlight also considered standards issued by the Council for Basic Education (1994) in *A Vision for Learning,* which provides a definition of academic knowledge needed to succeed in education. Other resources include Roeber and Stemmper 1989, Levin and Rumberger 1988, Raizen 1989, and Wirth 1993.

*Lessons Learned: (1) Several excellent documents contain desired workplace and academic performances. Interpersonal and intrapersonal expectations may be more difficult to reference. (2) Community involvement in the development process may help to define locally valued character attributes. (3) A tolerance for ambiguity is necessary. Trying to predict our children's future through the filter of our present has all the assurances of crystal ball gazing.*

### Holistic Learner Outcomes

Once Starlight staff members analyzed future life performances and expectations, they felt ready to revisit holistic learner outcomes, this time equipped with a basis on which to brainstorm. The going was still rough. Not unlike the experience of many other schools, Starlight's brainstormed list ranged from outcomes that bore a striking resemblance to skill-based objectives to bigger outcomes statements such as "be a global citizen." Staff realized they needed a definition of outcomes on which they could all agree. They detoured again to the literature and found the following helpful guideposts defining outcomes:

- What students should know or be able to do to function effectively in the 21st century (Martin-Kniep and Kniep 1992)
- A culminating demonstration of learning (Brandt 1992/1993)
- Upon completion of their high school studies, what our students should know, be able to do, feel, and believe (Fitzpatrick 1991)

Although Fitzpatrick's definition is becoming a widely accepted standard for defining outcomes, Flynn (1992) pushes the paradigm a little farther. He defines outcomes as "what a student should understand and be able to create." Starlight staff members agreed that Flynn's definition best matched what they had in mind and accepted it as their own. However, the definition alone was insufficient to address the dilemma of the range of items and overwhelming number of outcomes generated in brainstorming. Once again, the road became so bumpy and uncomfortable that another detour was clearly indicated.

This time, the side trip led to a set of criteria against which the outcomes on the brainstormed list could be evaluated. Starlight defined their criteria for holistic outcomes as follows:

Holistic outcomes are principle-based and research supported, assessable, analyzable, observable, and exhibit these characteristics:

- Having holistic application to all members of the community of learners,
- Describing the whole person,
- Reflecting personal and public value,
- Congruence with the values and beliefs of the community of learners,
- Present and future relevance,
- Being dynamic,
- Being robust enough to accommodate enhanced meaning over time, and
- Having transformative power

With these criteria in hand, the work of constructing a set of holistic student outcomes progressed smoothly and resulted in a set of four learner outcomes, presented in Figure 2. Starlight's road had been made smooth by their prior research into documents such as the SCANS report, and their examination of examples of outcomes from other districts and schools. This work supplied them with a basis upon which to justify their outcomes, and it reinforced the norm of making decisions based on research and data.

The definition of outcomes was a profound event for the Starlight School community. The process and product established a new paradigm for structuring the practices and experiences at the school. The act of reaching consensus on an initial set of student outcomes created a sense of mild euphoria.

FIGURE 2

### STARLIGHT ELEMENTARY SCHOOL LEARNER OUTCOMES

All Starlight learners will be:

**Powerful Thinkers**

Learners will utilize critical and creative thinking skills in problem solving and artistic expression, both independently and collaboratively. Learners will apply academic skills in making decisions that affect their lives now and in the future.

**Responsible Citizens**

Learners will value and participate in a democratic and culturally diverse society. Learners will be competent in using effective interpersonal and social skills. Learners will sustain mutually supportive relationships. Learners will effectively interact within society and will accept responsibility for self and their relationship to the environment.

**Self-Confident Individuals**

Learners will have a positive self-concept, exhibiting motivation, confidence ,and self-sufficiency. Learners will be flexible, adaptable individuals who make educated choices about behavior that supports physical, mental, emotional, and social well-being.

Predictably, the sense of well-being was short-lived as the realities of trying to achieve those outcomes set in. Tension was again on the rise as a plethora of questions emerged.

- What do we do with our outcomes?
- Do we have to reinvent everything we do to achieve these outcomes?
- How can we keep our outcomes from being a list of 200 skills objectives?
- How do we measure our outcomes?
- What is the difference between a performance indicator and an outcome?
- Do rubrics define performances or outcomes?
- How can we link outcomes to performances?
- How can we bundle outcomes (or indicators) into performances?
- What constitutes a profound performance worthy of being used for the demonstration of outcomes?
- What is the balance between being adequately general and specific enough?
- How many rubrics do we need to measure student achievement?
- How do we keep a rubric from becoming a floor or a ceiling?

Our experience with a variety of schools suggests that when questions of this nature are asked and their implications considered, real fear (which manifests in overt and covert resistance) is experienced. Restructuring of this magnitude gets to the core of people's senses of personal competence and willingness to take risks. Major roadblocks are introduced (no time, the district won't let us, parents won't understand, who is going to do all this, there aren't any good models, . . . ) as defenses and anxiety reach an all-time high. True leadership skill is necessary at this point lest the roadblocks result in turning the car around and going back to the safety of where the journey began. Keeping confidence and spirits high requires the leader to be cheerleader, coach, and guardian of the covenant of shared values (Sergiovanni 1992).

Starlight's staff gained reassurance when it considered the work its students were already doing. The teachers found that the students were currently making substantial progress toward the outcomes. Besides being reassuring, the examination of day-to-day student work unlocks answers to questions posed at this point in development. The top-down paradigm would have us move from the outcomes directly to lists of specific performances or objectives. Instead, if we move to the domain of student work, we begin to see the outcomes being performed in context, in a more authentic manner.

*Lessons Learned: (1) The term holistic embodies two paradigm shifts in the conceptualization of student work. First, holistic defines learning in terms of the whole person in life context. Second, holistic outcomes transcend place and time. They are robust enough to accommodate our best thinking on relevant learning for all ages, all peoples, and all times. Defined in this manner, holistic learner outcomes are performing serious work in the reconceptualization of learner success. (2) Criteria for what constitutes an outcome are helpful in the development of outcomes. (3) The frustrations and anxiety of the process should be viewed as productive and necessary, and as personal examples of what learners feel when they engage in constructivist activities.*

### Day-to-Day Performances

Day-to-day performances are those activities currently being performed by students in their school experience. As the Starlight community shifted its paradigm toward authentic achievement, outcomes became a useful lens through which to view student performances. As teachers observed students writing papers, working on science projects, and researching topics, they began to ask questions about how these activities inform the definition of the outcomes and how the outcomes inform an understanding of the performances. This investigative process was useful in refining the outcomes, building a consensus, and easing anxiety. As with all good learning, when teachers made connections between the known of current student work and the unknown nature of future work, the tasks to come seemed more manageable.

*Lessons Learned: (1) The process of seeking evidence of the outcomes in student work is much like the phenomenon that happens when a person is contemplating buying a new car. At the point when a person has focused on a particular automobile, he or she sees that car everywhere. So it is with outcomes. When teachers focus on particular outcomes, related behaviors and performances become very visible. (2) It is not imperative to arrive at consensus on outcomes at the initial brainstorming session. "Test driving" the outcomes in the context of day-to-day student performances to assess their power, potential, and robustness is valuable. After school community members have gathered experience with tentative outcomes, consensus building becomes an easier task.*

### Performance Indicators

Performance indicators should define the critical dimensions or attributes of an outcome essential for meaningful achievement. They show an audience that students are achieving the outcome (Johnson 1992). Starlight's reflections on students' day-to-day performances led them to conclude that indicators had to be:

- observable expressions;
- evidence of students' understanding, knowing, and ability to create new knowledge;
- measurable;
- clearly defined with minimal ambiguity; and
- consistent and related to the outcome.

Figure 3 lists Starlight School's performance indicators.

FIGURE 3

## EXAMPLES OF PERFORMANCE INDICATORS AT STARLIGHT SCHOOL

In order to become effective, literate communicators, Starlight students will

- Verbally express ideas, feelings, and opinions on a topic.
- Actively listen and respond to another person's verbal expressions.
- Show interest in reading by choosing a book to read in free time, expressing a desire to read, and enjoying story time.
- Express ideas, feelings, and opinions about a topic in writing through various forms (narrative, descriptive, expository, and experiential).
- Express self through physical movement and gestures in various forms (pantomime, sign language, facial expression, drama, role play, body language, and dance).
- Express self musically by singing, humming, clapping-snapping-tapping, demonstrating rhythm by hand and/or with an instrument, playing an instrument, or chanting and rapping.
- Use mathematical, logical thinking in problem solving, identifying patterns, solving puzzles and codes, and playing board games.

*Lesson Learned: A real danger at this point is that the process of outcome development will spiral off into a scope-and-sequence listing of 200 skill objectives, placing us right back where we began. To avoid this pitfall we must again focus on student work.*

### Authentic Performances

Authentic performance is the link between holistic learner outcomes, authentic assessment, and integrated, meaning-centered curriculum. An authentic performance provides students with an opportunity to demonstrate understanding and create new knowledge. Expressions of specific skills and the knowledge necessary to arrive at the exhibition are embedded in authentic performances.

At this point, Starlight had much practice as a community of learners. When confronted with the challenge of constructing authentic performances, school members headed for the literature and found three criteria for defining performance as authentic achievement (Newmann and Wehlage 1993, Newmann and Archbald 1992, Archbald and Newmann 1992). They found that authentic achievement is a vision of education that asks the learner to manipulate knowledge to produce some form of new discourse or performance, or some type of product with personal meaning and public value. Specifically, authentic achievement occurs when the following three events take place:

(1) Students construct meaning and produce new knowledge.

(2) Students use disciplined inquiry to construct meaning.

(3) Students aim their work toward production of discourse, products, and performances that have value or meaning beyond success in school.

Newmann and Archbald (1992) detail examples of authentic performances that have aesthetic, utilitarian, or personal value such as writing letters, news articles, insurance claims, or poems; speaking a foreign language; creating a painting or piece of music; building; producing products; or communicating ideas. Other examples of authentic performances include the production of a play, the conducting of conflict mediation, the creation of a business plan, the design of a science experiment or community service project, and the holding of a debate over a compelling local or national issue. Wiggins' writings (1989) and videotapes also provide rich examples of student projects and exhibitions. Starlight staff learned that the goal is to make assessment tasks like the real-life challenges a person would typically face as a citizen, writer, businessperson, scientist, community leader, or historian.

*Lesson Learned: Outcomes more easily define authentic performances than isolated skill and knowledge acquisition. Teachers can design authentic student work relatively easily when they apply the criteria for authentic achievement and have access to excellent models of authentic student work.*

## Rubrics

A rubric provides a description of various, qualitative levels of performance. This standard bearer and assessment tool develops from the interplay of the performance indicators and the authentic performance. A well-integrated authentic performance will cover several performance indicators. A single, authentic performance might have several rubrics addressing specific elements. For example, an oral history might have one rubric for research, one for writing, one for history, and one for oral presentation.

The design of a rubric is achieved through "backward mapping" (McDonald 1992) from a vision of what student work might look like in its final form (Johnson 1992). The rubric should present the student with a picture of what the final performance will look like as an authentic, contextualized whole in terms of content, skills, and behaviors. Those being assessed by a rubric (most often students) should be involved in its design. For example, one class of special education students learning to communicate through the production of photojournalism essays designed a rubric by examining the work of photojournalists. This examination resulted in a description of the knowledge, skills, and behaviors needed to produce a photojournalistic essay.

When operationalizing their holistic outcomes, Starlight School staff members focused first on the outcome about being a "responsible citizen." Working toward the achievement of this outcome led them to investigate and implement the concept of class meetings in which the students participate in a student-centered agenda. As the staff gained experience with class meetings (the day-to-day authentic performances), they could identify the knowledge, skills, and behaviors that constitute a class meeting. These observations formed the substance for creating the rubric.

One observation in particular triggered a new insight into the design of rubrics. A kindergarten teacher recounted how, after six months of class meetings, a student approached her with a request to discuss an issue at the class meeting. This was the first time any child had made such a request. This event defined a behavior not originally described by the rubric, but valued enough to be added: Does a child understand what types of issues are appropriate for class meetings, and does the child have the skills to introduce and explain such an issue? Furthermore, the staff at Starlight realized that the rubric itself must be intelligible to the learner. Consequently, the possible forms a rubric may take must often go beyond a written description to include pictures, videotape, audiotapes, and objects.

*Lessons Learned: (1) The rubric is a joint creation of students and teachers. (2) Student work informs rubric development in an ongoing way. (3) The rubric and exemplars are the starting point of learning. (4) The rubric must be communicated in terms students understand.*

## Experimental Performances and Standards

The establishment of rubrics is only the first step in a cycle of verification and validation. First, a period of experimentation is needed to validate the rubric scales relative to actual student work. Second, data from actual student work are required to determine

FIGURE 4

## RUBRIC EXAMPLE
## OUTCOME: RESPONSIBLE CITIZEN
## AUTHENTIC PERFORMANCE CONTEXT: CLASS MEETINGS

**Team Leader**

- Brings important ideas to the group
- Is careful not to hurt others' feelings
- Listens to others' ideas
- Works hard to involve everyone
- Builds on others' ideas
- Supports team decisions

5

**Team Member**

- Is careful not to hurt others' feelings
- Listens to others' ideas
- Builds on others' ideas
- Supports team decisions

4

**Team Participant**

- Listens to others' ideas
- Repeats the ideas of others
- Supports team decisions

3

**Team Observer**

- Listens to others' ideas

2

**New Member**

- Not engaged in meeting

1

the standards or benchmarks on the rubric scale. As staff members determine authentic performances and develop rubrics they find it useful to pilot those authentic performances. Examination of the resulting student work further informs the clarity and completeness of the rubric. The experimental performances enable the staff to establish: the validity of the rubric, defined by Case (1992) as the closeness of "fit between the kinds of attributes actually measured by the assessment device and those educational goals that we value"; its ability to reliably and validly measure the real, actual, or genuine thing as opposed to measuring a poor substitute (Case 1992); and its ability to accommodate a full range of student learning styles, rates, and intelligences, and enhance learning. Figure 4 shows an example of a rubric.

The second purpose of experimental performances is the production of examples of student work for use in defining levels of quality for the rubric. A standard on a rubric is an exemplary performance that serves as a benchmark (Wiggins 1991). Identifying two types of standards in a rubric is useful. The first standard is that of student work. What are the exemplars produced by students? A second standard is the best exemplars from humankind. What are the "bests" that humans have been able to achieve as musicians, composers, photographers, and so on.

*Lessons Learned: (1) Recognition of the need for experimental performances greatly reduces the staff's stress level, and provides essential data necessary to develop valid rubrics and powerful learning experiences. (2) Students are not intimidated or discouraged by high standards and their exemplars any more than a 10-year-old basketball player is intimidated by watching the NBA.*

## Starlight's Work in Progress

The Starlight journey is far from over. As staff members design curriculum, constructing authentic student work, they are confronting three big questions:

(1) How do we transition to a comprehensive curriculum of profound performances that meet the highest standards of authenticity?

(2) How do we manage the time and resource requirements of rubric-based assessments?

(3) How do we report student achievement and provide meaningful feedback to our students, parents, community, and to ourselves?

These questions serve as guideposts directing the next legs of this journey, a journey that will no doubt offer up a fair measure of agony and ecstasy.

## REFERENCES

Archbald, D.A., and F.M. Newmann. (1992). "Approaches to Assessing Authentic Achievement." *Toward a New Science of Educational Testing and Assessment*. Albany, N.Y.: SUNY Press.

Brandt, R. (December 1992/January 1993). "On Outcome-Based Education: A Conversation with Bill Spady." *Educational Leadership* 50, 4: 66-70.

Case, R. (1992). "On the Need to Assess Authentically." *Holistic Education Review* 5, 4: 14-23.

Council for Basic Education. (1994). *A Vision for Learning*. Washington, D.C.: Author.

Flynn P. (1992). "The Standard Is Understanding, the Assessment Criterion Is Creation." *Authentic Assessment*. Albany: New York State United Teachers.

Fitzpatrick, K.A. (May 1991). "Restructuring to Achieve Outcomes of Significance for All Students." *Educational Leadership* 48, 8: 18-22.

Johnson, B. (1992). "Creating Performance Assessments." *Holistic Education Review* 5, 4: 38-44.

Levin, H.M., and R.W. Rumberger. (November 1988). "Education, Work and Employment: Present Issues and Future Challenges in Developed Countries." XXVth Anniversary Conference of the International Institute for Educational Planning.

Martin-Kniep, G.O., and W.M. Kniep. (1992). "Alternative Assessment: Essential, Not Sufficient for Systemic Change." *Holistic Education Review* 5, 4: 4-13.

McDonald, J.P. (1992) "Dilemmas of Planning Backwards: Rescuing a Good Idea." *Studies on Exhibitions No. 3*. Providence, R.I.: Coalition of Essential Schools.

Newmann, F.M., and D.A. Archbald. (1992). "The Nature of Authentic Academic Achievement." *Toward a New Science of Educational Testing and Assessment*. Albany, N.Y.: SUNY Press.

Newmann, F.M., and G.G. Wehlage. (April 1993). "Five Standards of Authentic Instruction." *Educational Leadership* 50, 7: 8-12.

Roeber, E., W. Brown, and P. Stemmper. (June 1989). "Employability Skills: Defining a New Purpose for Public Schools." Lansing, Mich.: Michigan Department of Education.

Raizen, S. (December 1989). "Reforming Education for Work: Cognitive Science Perspective." (ED 314 642) Berkeley, Calif.: National Center for Research in Vocational Education, University of California.

Sergiovanni, T.J. (1992). *Moral Leadership*. San Francisco: Jossey-Bass.

U.S. Department of Labor, Secretary's Commission on Achieving Necessary Skills. (April 1992). *Learning a Living: A Blueprint for High Performance. A SCANS Report for America 2000*. Washington, D.C.: Author.

Wiggins, G. (May 1989). "A True Test: Toward More Authentic Assessment and Equitable Assessment." *Phi Delta Kappan* 70, 9: 703-713.

Wiggins, G. (February 1991). "Standards, Not Standardization: Evoking Quality Student Work." *Educational Leadership* 48, 5: 18-25.

Wirth, A. G. (January 1993). "Education and Work: The Choices We Face." *Phi Delta Kappan* 74, 5: 360-366.

# III-20

# The Discipline of Building Shared Vision

PETER M. SENGE
INNOVATION ASSOCIATES

## Encouraging Personal Vision

Shared visions emerge from personal visions. This is how they derive their energy and how they foster commitment. As Bill O'Brien of Hanover Insurance observes, "My vision is not what's important to you. The only vision that motivates you is your vision." It is not that people care only about their personal self-interest—in fact, people's personal visions usually include dimensions that concern family, organization, community, and even the world. Rather, O'Brien is stressing that caring is *personal*. It is rooted in an individual's own set of values, concerns, and aspirations. This is why genuine caring about a shared vision is rooted in personal visions. This simple truth is lost on many leaders, who decide that their organization must develop a vision by tomorrow!

Organizations intent on building shared visions continually encourage members to develop their personal visions. If people don't have their own vision, all they can do is "sign up" for someone else's. The result is compliance, never commitment. On the other hand, people with a strong sense of personal direction can join together to create a powerful synergy toward what I/we truly want.

Personal mastery is the bedrock for developing shared visions. This means not only personal vision, but commitment to the truth and creative tension—the hallmarks of personal mastery. Shared vision can generate levels of creative tension that go far beyond individuals' "comfort levels." Those who will contribute the most toward realizing a lofty vision will be those who can "hold" this creative tension: remain clear on the vision and continue to inquire into current reality. They will be the ones who believe deeply in their ability to create their future, because that is what they experience personally.

In encouraging personal vision, organizations must be careful not to infringe on individual freedoms. As was discussed in chapter 9, "Personal Mastery," no one can give another "his vision," nor even force him to develop a vision. However, there are positive actions that can be taken to create a climate that encourages personal vision. The most direct is for leaders who have a sense of vision to communicate that in such a way that others are encouraged to share their visions. This is the art of visionary leadership—how shared visions are built from personal visions.

---

Reprinted with permission from *The Fifth Discipline: The Art and Practice of the Learning Organization.* (New York: Doubleday/Currency, 1990), pp. 211-226. Copyright © 1990 by Peter M. Senge.

## From Personal Visions to Shared Visions

How do individual visions join to create shared visions" A useful metaphor is the hologram, the three-dimensional image created by interacting light sources.

If you cut a photograph in half, each part shows only part of the whole image. But if you divide a hologram, each part shows the whole image intact. Similarly, as you continue to divide up the hologram, no matter how small the divisions, each piece still shows the whole image. Likewise, when a group of people come to share a vision for an organization, each person sees his own picture of the organization at its best. Each shares responsibility for the whole, not just for his piece. But the component "pieces" of the hologram are not identical. Each represents the whole image from a different point of view. It's as if you were to look through holes poked in a window shade; each hole would offer a unique angle for viewing the whole image. So, too, is each individual's vision of the whole unique. We each have our own way of seeing the larger vision.

When you add up the pieces of a hologram, the image of the whole does not change fundamentally. After all, it was there in each piece. Rather the image becomes more intense, more lifelike. When more people come to share a common vision, the vision may not change fundamentally. But it becomes more alive, more real in the sense of a mental reality that people can truly imagine achieving. They now have partners, "cocreators"; the vision no longer rests on their shoulders alone. Early on, when they are nurturing an individual vision, people may say it is "my vision." But as the shared vision develops, it becomes both "my vision" and "our vision."

The first step in mastering the discipline of building shared visions is to give up traditional notions that visions are always announced from "on high" or come from an organization's institutionalized planning processes.

In the traditional hierarchical organization, no one questioned that the vision emanated from the top. Often, the big picture guiding the firm wasn't even shared—all people needed to know were their "marching orders," so that they could carry out their tasks in support of the larger vision. Ed Simon of Herman Miller says, "If I was the president of a traditional authoritarian organization and I had a new vision, the task would be much simpler than we face today. Most people in the organization wouldn't need to understand the vision. People would simply need to know what was expected of them."

That traditional "top-down" vision is not much different from a process that has become popular in recent years. Top management goes off to write its "vision statement," often with the help of consultants. This may be done to solve the problem of low morale or lack of strategic direction. Sometimes the process is primarily reflective. Sometimes it incorporates extensive analysis of a firm's competitors, market setting, and organizational strengths and weaknesses. Regardless, the results are often disappointing for several reasons.

First, such a vision is often a "one-shot" vision, a single effort at providing overarching direction and meaning to the firm's strategy. Once it's written, management assumes that they have now discharged their visionary duties. Recently, one of my Innovation Associates colleagues was explaining to two managers how our group works with vision. Before he could get far, one of the managers interrupted. "We've done that," he said. "We've already written our vision statement." "That's very interesting," my colleague responded. "What did you come up with?" The one manager turned to the other and asked, "Joe, where is that vision statement anyhow?" Writing a vision statement can be a first step in building shared vision but, alone, it rarely makes a vision "come alive" within an organization.

The second problem with top management going off to write their vision statement is that the resulting vision does not build on people's personal visions. Often, personal visions are ignored altogether in the search for a "strategic vision." Or the "official vision" reflects only the personal vision of one or two people. There is little opportunity for inquiry and testing at every level so that people feel they understand and own the vision. As a result, the new official vision also fails to foster energy and commitment. It simply does not inspire people. In fact, sometimes, it even generates little passion among the top management team who created it.

Lastly, vision is not a "solution to a problem." If it is seen in that light, when the "problem" of low morale or unclear strategic direction goes away, the energy behind the vision will go away also. Building shared vision must be seen as a central element of the daily work of leaders. It is ongoing and never-ending. It is actually part of a larger leadership activity: designing and nurturing what Hanover's Bill O'Brien calls the "governing ideas" of the enterprise—not only its vision per se, but its purpose and core values as well. As O'Brien says, "The governing ideas are far more important and enduring than the reporting chart and the divisional structure that so often preoccupy CEOs."

Sometimes, managers expect shared visions to emerge from a firm's strategic planning process. But for all the same reasons that most "top-down" visioning processes fail, most strategic planning also fails to nurture genuine vision. According to Hamel and Prahalad:

> Creative strategies seldom emerge from the annual planning ritual. The starting point

> for next year's strategy is almost always this year's strategy. Improvements are incremental. The company sticks to the segments and territories it knows, even though the real opportunities may be elsewhere. The impetus for Canon's pioneering entry into the personal copier business came from an overseas sales subsidiary—not from planners in Japan.[1]

This is not to say that visions cannot emanate from the top. Often, they do. But sometimes they emanate from personal visions of individuals who are not in positions of authority. Sometimes they just "bubble up" from people interacting at many levels. The origin of the vision is much less important than the process whereby it comes to be shared. It is not truly a "shared vision" until it connects with the personal visions of people throughout the organization.

For those in leadership positions, what is most important is to remember that their visions are still personal visions. Just because they occupy a position of leadership does not mean that their personal visions are *automatically* "the organization's vision." When I hear leaders say "our vision" and I know they are really describing "my vision," I recall Mark Twain's words that the official "we" should be reserved for "kings and people with tapeworm."

Ultimately, leaders intent on building shared visions must be willing to continually share their personal visions. They must also be prepared to ask, "Will you follow me?" This can be difficult. For a person who has been setting goals all through his career and simply announcing them, asking for support can make him feel very vulnerable.

John Kryster was the president of a large division of a leading home products company who had a vision that his division should be preeminent in its industry. This vision required not only excellent products but that the company supply the product to their "customer" (retail grocers), in a more efficient and effective manner than anyone else. He envisioned a unique worldwide distribution system that would get product to the customer in half the time and with a fraction of the cost in wastage and reshipments. He began to talk with other managers, with production workers, with distribution people, with grocers. Everyone seemed enthusiastic, but pointed up that many of his ideas could not be achieved because they contradicted so many traditional policies of the corporate parent.

In particular, Kryster needed the support of the head of product distribution, Harriet Sullivan, who—while technically Kryster's peer in the firm's matrix organization—had fifteen years more experience. Kryster prepared an elaborate presentation for Sullivan to show her the merits of his new distribution ideas. But for every piece of supporting data he offered, Sullivan had a countering criticism. Kryster left the meeting thinking that the doubters were probably right.

Then he conceived of a way to test the new system out in only one geographic market. The risk would be less, and he could gain the support of the local grocery chain which had been especially enthusiastic about the concept. But what should he do about Harriet Sullivan? His instincts were just not to tell her. After all, he had the authority to undertake the experiment himself, using his own distribution people. Yet, he also valued Sullivan's experience and judgment.

After a week of mulling it over, Kryster went back to ask for Sullivan's support. This time, though, he left his charts and data at home. He just told her why he believed in the idea, how it could forge a new partnership with customers, and how its merits could be tested with low risk. To his surprise, the crusty distribution chief started to offer help in designing the experiment. "When you came to me last week," she said, "you were trying to convince me. Now, you're willing to test your idea. I still think it's wrongheaded, but I can see you care a great deal. So, who knows, maybe we'll learn something."

That was five years ago. Today, John Kryster's innovative distribution system is used worldwide by almost all the corporation's divisions. It has significantly reduced costs and been part of broad strategic alliances the corporation is learning to forge with retail chains.

When visions start in the middle of an organization the process of sharing and listening is essentially the same as when they originate at the top. But it may take longer, especially if the vision has implications for the entire organization.

Bart Bolton was a middle manager in IS (Information Systems) at Digital Equipment Corporation when, back in 1981, he and a small group of colleagues began to form an idea of Digital as an interconnected organization. "A group of us had been together at a workshop, and when we came back we just started talking about how we were going to turn around IS. The fundamental problem as we all saw it was that there simply was no IS vision. Everyone argued about the 'how to's' but no one knew the 'what.' Yet, we felt we could see an end result that was really worth going for. We didn't know exactly what it would look like, but the idea of tying the organization together electronically just felt 'right.' Given our products and technology we could become one of the first, if not the first large corporation that was totally and completely electronically interconnected." The idea was so exciting that he couldn't sleep much for several days as he thought about the implications.

But in 1981, no one had any idea how this could be done. "It was simply beyond the realm of what was possible at that time. We could transfer files between computers, but we couldn't network. There was some networking software under development but there were lots of problems with it. Perhaps, if we worked really hard at it we could interconnect ten or twenty machines, but no one even dreamed of interconnecting a hundred machines, let alone thousands. Looking back, it was like they say about Kennedy when he announced the 'Man on the Moon' vision—we knew about 15 percent of what we needed to know to get there. But we knew it was right."

Bolton and his compatriots had no "authority" to pursue the idea, but they couldn't stop thinking about it. In November 1981, he wrote a short paper which he read to all the senior IS people at a staff meeting. In it he said that the organization of the future would involve new IS technologies, would see "data as a resource just like the organization of the past saw capital and people as resources," and that "networks would tie together all the functions." "When I finished, no one spoke. It was like being in church. I really thought I'd blown it. My boss, Al Crawford, the head of IS, suggested a ten-minute break. When people came back, all they wanted to know was, 'How do we promote it? How can we make it happen?' My only response was, 'This has got to be your vision not mine, or it will never happen.'"

"I knew the guys at the top had to be 'enrolled,' and my job was to help them lead. By enrolling others, they too would become messengers." An IS group prepared a 35-mm slide show to be used by Crawford throughout the organization. He came up with the image of "wiring up the corporation." "It became incredibly exciting," says Bolton, "to watch the vision build, each person adding something new, refining it and making it come alive. We literally began talking about the 'copper wires running around the world.' "

Crawford presented the slide show to all Digital's major functional staffs in 1982. The idea, "the what," started to take hold. Then the IS organization created five overlapping programs to tackle the "how to's": a network program, a data program, an office automation program, a facilities program, and an applications program. By 1985 the first network was in place. By 1987, over 10,000 computers were on line. Today, Digital has over 600 facilities in over 50 countries and they are all interconnected. There are over 43,000 computers interconnected. Digital is now seen by experts as one of the pioneer "networked organizations." Moreover, the "networked organization" is a dominant theme in Digital's marketing strategy and advertising.

Organizational consultant Charlie Kiefer says that, "Despite the excitement that a vision generates, the process of building shared vision is not always glamorous. Managers who are skilled at building shared visions talk about the process in ordinary terms. 'Talking about our vision' just gets woven into day-to-day life. Most artists don't get very excited about the *process of* creating art. They get excited about the results." Or, as Bill O'Brien puts it, "Being a visionary leader is not about giving speeches and inspiring the troops. How I spend my day is pretty much the same as how any executive spends his day. Being a visionary leader is about solving day-to-day problems with my vision in mind."

Visions that are truly shared take time to emerge. They grow as a by-product of interactions of individual visions. Experience suggests that visions that are genuinely shared require ongoing conversation where individuals not only feel free to express their dreams, but learn how to listen to each others' dreams. Out of this listening, new insights into what is possible gradually emerge.

Listening is often more difficult than talking, especially for strong-willed managers with definite ideas of what is needed. It requires extraordinary openness and willingness to entertain a diversity of ideas. This does not imply that we must sacrifice our vision "for the larger cause." Rather, we must allow multiple visions to coexist, listening for the right course of action that transcends and unifies all our individual visions. As one highly successful CEO expressed it: "My job, fundamentally, is listening to what the organization is trying to say, and then making sure that it is forcefully articulated."

## Spreading Visions: Enrollment, Commitment, and Compliance[2]

Few subjects are closer to the heart of contemporary managers than commitment. Prodded by studies showing that most American workers acknowledge low levels of commitment[3] and by tales of foreign competitors' committed work forces, managers have turned to "management by commitment," "high commitment work systems," and other approaches. Yet, real commitment is still rare in today's organizations. It is our experience that, 90 percent of the time, what passes for commitment is compliance.

Today, it is common to hear managers talk of getting people to "buy into" the vision. For many, I fear, this suggests a sales process, where I sell and you buy. Yet, there is a world of difference between "selling" and "enrolling." "Selling" generally means getting someone to do something that he might not do if they were in full possession of all the facts. "Enrolling," by contrast, literally means "placing one's name on the roll." Enrollment implies free choice, while "being sold" often does not.

"Enrollment is the process," in Kiefer's words, "of becoming part of something by choice." "Committed" describes a state of being not only enrolled but feeling fully responsible for making the vision happen. I can be thoroughly enrolled in your vision. I can genuinely want it to occur. Yet, it is still your vision. I will take actions as need arises, but I do not spend my waking hours looking for what to do next.

For example, people are often enrolled in social causes out of genuine desire, for example, to see particular inequities righted. Once a year they might make a donation to help in a fund-raising campaign. But when they are committed, the "cause" can count on them. They will do whatever it takes to make the vision real. The vision is pulling them to action. Some use the term "being source" to describe the unique energy that committed people bring toward creating a vision.

In most contemporary organizations, there are relatively few people enrolled—and even fewer committed. The great majority of people are in a state of "compliances." "Compliant" followers go along with a vision. They do what is expected of them. They support the vision, to some degree. But, they are not truly enrolled or committed.

Compliance is often confused with enrollment and commitment. In part, this occurs because compliance has prevailed for so long in most organizations, we don't know how to recognize real commitment. It is also because there are several levels of compliance, some of which lead to behavior that looks a great deal like enrollment and commitment:

The speed limit is fifty-five in most states in the United States today. A person who was genuinely compliant would never drive more than fifty-five. A person formally compliant could drive sixty to sixty-five because in most states you will not get a ticket so long as you are below sixty-five. Someone grudgingly compliant would stay below sixty-five and complain continually about it. A noncompliant driver would "floor it" and do everything possible to evade troopers. On the other hand, a person who was genuinely committed to a fifty-five mph speed limit would drive that speed even if it were not the legal limit.

### POSSIBLE ATTITUDES TOWARD A VISION

**Commitment:** Wants it. Will make it happen. Creates whatever "laws" (structures) are needed.

**Enrollment:** Wants it. Will do whatever can be done within the "spirit of the law."

**Genuine compliance:** Sees the benefits of the vision. Does everything expected and more. Follows the "letter of the law." "Good soldiers."

**Formal compliance:** On the whole, sees the benefits of the vision. Does what's expected and no more. "Pretty good soldier."

**Grudging compliance:** Does not see the benefits of the vision. But, also, does not want to lose job. Does enough of what's expected because he has to, but also lets it be known that he is not really on board.

**Noncompliance:** Does not see benefits of vision and will not do what's expected. "I won't do it; you can't make me."

**Apathy:** Neither for nor against vision. No interest. No energy. "Is it five o'clock yet?"

In most organizations, most people are in states of formal or genuine compliance with respect to the organization's goals and ground rules. They go along with "the program," sincerely trying to contribute. On the other hand, people in noncompliance or grudging compliance usually stand out. They are opposed to the goals or ground rules and let their opposition be known, either through inaction or (if they are grudgingly compliant) through "malicious obedience"—"I'll do it just to prove that it won't work." They may not speak out publicly against the organization's goals, but their views are known nonetheless (They often reserve their truest sentiments for the rest room or the cocktail lounge.)

Differences between the varying states of compliance can be subtle. Most problematic is the state of genuine compliance, which is often mistaken for enrollment or commitment. The prototypical "good soldier" of genuine compliance will do whatever is expected of him, willingly. "I believe in the people behind the vision; I'll do whatever is needed, and more, to the fullest of my ability." In his own mind, the person operating in genuine compliance often thinks of himself as committed. He is, in fact, committed, but only to being part of the team."

In fact, from his *behavior* on the job, it is often very difficult to distinguish someone who is genuinely compliant from someone who is enrolled or committed. An organization made up of genuinely compliant people would be light-years ahead of most organizations in productivity and cost effectiveness. People would not have to be told what to do more than once. They would be responsive. They would be upbeat and positive in their attitude and manner. They might also be a bit "drone-like," but not necessarily. If what was expected of high performers was to "take initiative" and be "proactive," they would exhibit those behaviors as well. In short, people in genuine compliance would do whatever they could to play by the "rules of the game," both the formal and subtle rules.

Yet, there *is* a world of difference between compliance and commitment. The committed person brings an energy, passion, and excitement that cannot be generated if you are only compliant, even genuinely compliant. The committed person doesn't play by the "rules of the game." He is responsible for the game. If the rules of the game stand in the way of achieving the vision, he will find ways to change the rules. A group of people truly committed to a common vision is an awesome force. They can accomplish the seemingly impossible.

Tracy Kidder, in his Pulitzer-prize-winning book *The Soul of a New Machine,* tells the story of a product development team at Data General, brought together by a talented team leader to create an ambitious new computer. Against a business atmosphere of urgency bordering on crisis, the team turned out a groundbreaking computer in remarkable time. Visiting with the team manager Tom West in the book, and team members several years later, I learned just how remarkable their feat was. They told me of a stage in their project where certain critical software was several months behind schedule. The three engineers responsible came into the office one evening and left the next morning. By all accounts they accomplished two to three months of work that evening—and no one could explain how. These are not the feats of compliance.

What then is the difference between being genuinely compliant and enrolled and committed? The answer is deceptively simple. People who are enrolled or committed truly *want* the vision. Genuinely compliant people accept the vision. They may want it in order to get something else—for example, to keep their job, or to make their boss happy, or to get a promotion. But they do not truly want the vision in and of itself. It is not their own vision (or, at least, they do not know that it is their own vision).

Highly desired, shared commitment to a vision can be an elusive goal. One executive VP at a consumer goods company deeply desired to turn the very traditional organization into a world-class competitor by developing shared commitment to a new business vision. But after a year's effort, people continued to follow orders and do what they were told.

At this point he began to see the depth of the problem. People in his organization had *never been asked to commit to anything in their careers*. All they had ever been asked to do was be compliant. That was all they knew how to do. That was their only mental model. No matter what he said about developing a real vision, about being truly committed, it didn't matter because they heard it within their model of compliance.

Once he grasped this, he shifted tactics. He asked, "What might people be able to commit to?" He initiated a "wellness program," reasoning if there was anything to which people might become committed, it would be their own health. Over time, some did. They began to see that true commitment was possible in the workplace, and a near "ear" for the vision was opened.

Traditional organizations did not care about enrollment and commitment. The command and control hierarchy required only compliance. Still, today, many managers are justifiably wary of whether the energy released through commitment can be controlled and directed. So, we settle for compliance and content ourselves with moving people up the compliance ladder.

## Guidelines for Enrollment and Commitment

Enrollment is a natural process that springs from your genuine enthusiasm for a vision and your willingness to let others come to their own choice.

- *Be enrolled yourself.* There is no point attempting to encourage another to be enrolled when you are not. That is "selling," not enrolling and will, at best, produce a form of superficial agreement and compliance. Worse, it will sow the seeds for future resentment.
- *Be on the level.* Don't inflate benefits or sweep problems under the rug. Describe the vision as simply and honestly as you can.
- *Let the other person choose.* You don't have to "convince" another of the benefits of a vision. In fact, efforts you might make to persuade him to "become enrolled" will be seen as manipulative and actually preclude enrollment. The more willing you are for him to make a free choice, the freer he will feel. This can be especially difficult with subordinates, who are often conditioned to feel as though they must go along. But you can still help by creating the time and safety for them to develop their own sense of vision.

There are many times when managers need compliance. They may want enrollment or commitment, but cannot accept anything below formal compliance. If that is the case, I recommend that you be on the level about it: "I know you may not agree wholeheartedly with the new direction, but at this juncture it is where the management team is committed to heading. I need your support to help it happen." Being open about the need for compliance removes hypocrisy. It also makes it easier for people to come to their choices, which may, over time, include enrollment.

The hardest lesson for many managers to face is that, ultimately, *there is nothing you can do to get another person to enroll or commit.* Enrollment and commitment require freedom of choice. The guidelines above simply establish conditions most favorable to enrollment, but they do not *cause* enrollment. Commitment likewise is very personal; efforts to force it will, at best, foster compliance.

## Anchoring Vision in a Set of Governing Ideas

Building shared vision is actually only one piece of a larger activity: developing the "governing ideas" for the enterprise, its vision, purpose or mission, and core values. A vision not consistent with values that people live by day by day will not only fail to inspire genuine enthusiasm, it will often foster outright cynicism.

These governing ideas answer three critical questions: "What?" "Why?" and "How?"

- Vision is the "What?"—the picture of the future we seek to create.
- Purpose (or "mission") is the "Why?"—the organization's answer to the question, "Why do we exist?" Great organizations have a larger sense of purpose that transcends providing for the needs of shareholders and employees. They seek to contribute to the world in some unique way, to add a distinctive source of value.
- Core values answer the question "How do we want to act, consistent with our mission, along the path toward achieving our vision? "A company's values might include integrity, openness, honesty, freedom, equal opportunity, leanness, merit, or loyalty. They describe how the company wants life to be on a day-to-day basis, while pursuing the vision.

Taken as a unit, all three governing ideas answer the question, "What do we believe in?" When Matsushita employees recite the company creed: "To recognize our responsibilities as industrialists, to foster progress, to promote the general welfare of society, and to devote ourselves to the further development of world culture," they're describing the company *purpose*. When they sing the company song, about "sending our goods to the people of the world, endlessly and continuously, like water gushing from a fountain," they're proclaiming the corporate *vision*. And when they go to inhouse training programs that cover such topics as "fairness," "harmony and cooperation," "struggle for betterment," "courtesy and humility," and "gratitude," the employees are learning the company's deliberately constructed values. (Matsushita, in fact, calls them its "spiritual values.")[4]

At Hanover Insurance, articulating all three of these "governing ideas" made an enormous difference in the firm's revival from near bankruptcy to a leader in the property and liability industry. Hanover's experience also illustrates the interdependencies among vision, values, and purpose.

"Early on," says O'Brien, "we recognized that there is a burning need for people to feel part of an ennobling mission. If it is absent many will seek fulfillment only in outside interests instead of in their work.

"But we also discovered that stating a mission or purpose in words was not enough. It ends up sounding like 'apple pie and motherhood.' People need visions to make the purpose more concrete and tangible. We had to learn to 'paint pictures' of the type of organization we wanted to be. My simple vision for the company is 'unquestioned superiority.' This simple term has great meaning for me. It leads me to envision an organization that serves the customer in unique ways, maintains a reputation for quality and responsibility, and creates a unique environment for its employees.

"Core values are necessary to help people with day-to-day decision making. Purpose is very abstract. Vision is long term. People need 'guiding stars' to navigate and make decisions day to day. But core values are only helpful if they can be translated into concrete behaviors. For example, one of our core values is 'openness,' which we worked long and hard to understand—finally recognizing that it requires the skills of reflection and inquiry within an overall context of trusting and supporting one another."

## Positive Versus Negative Vision

"What do we want?" is different from "What do we want to avoid?" This seems obvious, but in fact negative visions are probably more common than positive visions. Many organizations truly pull together only when their survival is threatened. They focus on avoiding what people don't want—being taken over, going bankrupt, losing jobs, not losing market share, having no downturns in earnings, or "not letting our competitors beat us to market with our next new product." Negative visions are, if anything, even more common in public leadership, where societies are continually bombarded with visions of "anti-drugs," "anti-smoking," "anti-war," or "anti-nuclear energy."

Negative visions are limiting for three reasons. First, energy that could build something new is diverted to "preventing" something we don't want to happen. Second, negative visions carry a subtle yet unmistakable message of powerlessness: our people really don't care. They can pull together only when there is sufficient threat. Lastly, negative visions are inevitably short term. The organization is motivated so long as the threat persists. Once it leaves, so does the organization's vision and energy.

There are two fundamental sources of energy that can motivate organizations: fear and aspiration. The power of fear underlies negative visions. The power of aspiration drives positive visions. Fear can produce extraordinary changes in short periods, but aspiration endures as a continuing source of learning and growth.

## Creative Tension and Commitment to the Truth

In Chapter 9 ("Personal Mastery"), I argued that personal vision, by itself, is not the key to more effective creativity. The key is "creative tension," the tension between vision and reality. The most effective people are those who can "hold" their vision while remaining committed to seeing current reality clearly.

This principle is no less true for organizations. The hallmark of a learning organization is not lovely visions floating in space, but a relentless willingness to examine "what is" in light of our vision.

IBM in the early 1960s, for example, carried out an extraordinary series of experiments in pursuit of a daring vision, a single family of computers that would make virtually all its previous machines obsolete. In the words of a *Fortune* writer, IBM staked "its treasure, its reputation, and its position of leadership in the computer field" on a radical new concept: a series of compatible machines serving the broadest possible range of applications, from the most sophisticated scientific applications to the relatively small business needs.[5]

Jay Forrester once remarked that the hallmark of a great organization is "how quickly bad news travels upward." IBM's capacity to recognize and learn from its mistakes proved pivotal during this period. One of the most discouraging was an early attempt at a high-end machine called "Stretch," introduced in 1960. IBM CEO Tom Watson, Jr., effectively killed the project in May 1961, after only a few had been sold. (Watson cut Stretch's hefty $13.5 million price tag almost in half, thereby making it uneconomical to produce.) To him, there was little choice: the machine did not satisfy its customers, never achieving more than 70 percent of its promised specifications. A few days later, Watson spoke candidly to an industry group. "Our greatest mistake in Stretch," he said, "is that we walked up to the plate and pointed at the center field stands. When we swung, it was not a homer but a hard line drive to the outfield. We're going to be a good deal more careful about what we promise in the future."

Indeed they were. Under the direction of many of the same men who had learned from Stretch, IBM introduced the System 360 three years later, which proved to be the platform for its extraordinary growth over the next ten years.

### NOTES

[1] G. Hamel and C.K. Prahalad in "Strategic Intent," *Harvard Business Review*, May-June, 1989.

[2] The ideas expressed in this section come from many hours of discussion with my colleagues at Innovation Associates, notably Charles Kiefer, Alain Gauthier, Charlotte Roberts, Rick Ross, and Bryan Smith.

[3] For example, the Daniel Yankelovich and John Immerwahr study in 1983, which found that only 25 percent of U.S. workers said they were working as hard as they could be. "Are U.S. workers lazy" by Joani Nelson-Horchler, *Industry Week*, June 10, 1985, 47.

[4] M. Moskowitz, *The Global Marketplace* (New York: Macmillan Publishing Company, 1987).

[5] "IBM's $5,000,000,000 Gamble." *Fortune*, September 1966, and "The Rocky Road to the Marketplace," *Fortune*, October 1966 (two-part article).

# III-21

## In Search of Community

ERNEST L. BOYER
THE CARNEGIE FOUNDATION FOR THE ADVANCEMENT OF TEACHING

I'm absolutely delighted to be invited to join you at this wonderful convocation, which celebrates the 50th Anniversary of the Association for Supervision and Curriculum Development. ASCD is one of the nation's distinguished and most effective educational associations. You have, throughout the years, focused on teaching and on learning. The goal of this Association has from the very first been to achieve excellence for all. This Association has for more than 50 years been a true community of learning, and I salute you—all of you assembled in this room—for a half century of dedicated and inspired service to the nation, and most especially to our children.

This morning, at the launching of this convention, I should wish to celebrate communities of learning in the classroom, communities of learning that extend into the neighborhoods and into the families, but my focus this morning will be to talk about coherence within the curriculum itself. And I should like to talk about an issue that has perplexed educators and philosophers and parents for at least 1,000 years.

The question I should like to ask at the beginning of this Conference is, just what is an educated person? Or more precisely, what should we be teaching students in our schools as we approach century 21? To put my remarks in larger context, I'd like to begin by telling you a story.

### Human Differences, Human Similarities

In 1972, I was sitting in my office in Albany, New York. It was a dreary Monday morning, which is kind of a redundancy I know, and to avoid the pressures of the day, I turned instinctively to the stack of third-class mail that I'd kept perched precariously on the corner of my desk to create the illusion of being very, very busy. It's an old administrative trick. On top of the heap was the student newspaper from Stanford University. The headlines grabbed my attention because they announced that the faculty at Stanford had reintroduced a required course in Western Civilization after having abolished all requirements just three years before. I also was intrigued to discover that the students at Stanford were mightily offended by this brash act, and, in the front-page editorial, they declared a required course is an illiberal act. Then the editors concluded with this blockbuster question: How dare they impose uniform standards on non-uniform people?

Well, at first I was amused, and then, I must tell, troubled by that statement. I was troubled by that statement. I was troubled that some of America's most gifted students, after 14 or more years of formal learning, still had not learned the simple truth that while we're non-uniform (of course we're non-uniform, just look to your left and to your right and say

---

Paper presented at the 48th Annual Conference of the Association for Supervision and Curriculum Development, Washington, D.C., March 1993. Reprinted with permission of the author.

a prayer of thanks) we still have many things in common. They had not discovered the essential truth that with all of our diversity—which I celebrate today—there are still characteristics at the very core of our existence that bind us all together. There are, to put it simply, connections to be found. This brings me to the central theme of my remarks this morning.

Being an educated person, in my opinion, surely means becoming well informed. It means developing our own aptitudes of interest. It means discovering the diversity that makes each of us unique and gives distinctiveness to our cultures; but there is another side to the equation, and it's called *connections*. Today, almost all students in the nation's schools complete their Carnegie units, they're handed a diploma; but I'm convinced that in our fragmented academic world what they often fail to gain is a more coherent view of knowledge and a more integrated, more authentic view of life. To put it simply, learning (for far too many years) has become an exercise in fragmentation, a pursuit of trivial pursuit.

And for far too long, education in this country has been based on seat-time, not on learning. It's enormously revealing that the most frequent question asked in many classrooms is, will we have this on the test? Frankly, I'm convinced the time has come to bury once and for all the old Carnegie unit. Further, since the foundation I now head created this academic measurement a century ago, I feel authorized this morning to officially declare the Carnegie unit obsolete.

I also am convinced that the proposed national assessment program should not be implemented until we are very clear about what schools should be teaching as we enter century 21. It's absolutely urgent that we not get the cart before the horse. But I return then to the question that brings me here this morning: What should we be teaching to our students? Just what does it mean to be an educated person?

## An Educated Person in Contemporary Society

To put it as simply as I can, an educated person is someone who is well informed, acts wisely, and continues learning; but to be truly educated means going beyond the isolated facts. It means putting learning in larger context. Above all, to return to my favorite word, it means discovering the "connectedness" of things. The Nobel Laureate, Barbara McClintock, wrote on one occasion, that everything is one. There is, she said, no way to draw a line between things. I wonder if Professor McClintock has looked at a school curriculum in recent years with all the separate academic boxes. Frank Cress, president of the National Academy of Sciences, said in a recent speech that scientists are in some respects like artists. To illustrate the point, he said that the magnificent double helix, which broke the genetic code, was not only irrational, it was beautiful as well. Yet, in most schools, science and art teachers live in absolutely separate worlds. They don't communicate with one another.

When I read Frank's speech, I thought of the early days of the space program. When I used to watch with fascination the lift-offs from Cape Kennedy, I was intrigued that in the final moments, as they went from 10 to 9 to 8 to 7, the scientists and engineers sat with anxiety on their faces; and then when they got to 3 to 2 to 1, contact, and the space shuttles went rocketing into orbit, the technologists and technicians wouldn't say, well our formulas worked again. They would say, almost in unison, the word "beautiful." It always intrigued me that they chose an aesthetic term to describe a technological achievement, which raises the fascinating question of where does the work of the scientist and artist begin and end? Is it possible that across the academic disciplines, which we consider so preciously to protect, that we're all inquiring, not just for intellectual searches, but down deep we're looking for patterns and relationships that bring intellectual satisfaction, too. Our scientists at their best are, in fact, artists.

The world-renowned physicist Victor Weisskopf was asked on one occasion, what gives you hope in troubled times? He replied, Mozart and quantum mechanics. But where in the school experience do students begin to see, even at the most fundamental level, connections such as these? Of course, students should become knowledgeable in history, and in literature, and in science, and all the rest; but what we urgently need today are students who can put their learning in larger context. For this to be accomplished, I believe we need a more thematic curriculum—a curriculum with more coherence, one that goes beyond these separate academic subjects and uses the disciplines—not as ends—but to illuminate larger, more integrative ends. Where do we begin?

## Core Commonalities

Several years ago,* in a book called *Quest for Common Learning*, I suggested as just one approach that we might organize the school curriculum, not on the basis of disciplines or departments, but on the basis of what might be called the "core commonalities." By core commonalties, I mean those universal human experiences that are shared by all people and all cultures on the planet. Those core experiences that make us truly human. While reflecting on the possibility of this thematic structure, I concluded that there are, in fact, eight human commonalities that bind us all together. Let me recite them for you very briefly here this morning.

## The Life Cycle

First, at the most basic level, we all share (and it's so obvious that we ignore it) the universal human experience of birth and growth and death. I've often thought if you want to make connections in another culture, take a baby. The life cycle binds us all together, and yet the sad truth is that most students go through life without reflecting on the mystery of their own existence. They complete 12 or even 16 years of formal schooling not considering the sacredness of their own bodies, not learning how to sustain wellness, and not pondering the imperative of death.

My wife is a certified nurse midwife and delivers babies, including seven grandchildren of her own; and Kay tells me of delivering the babies of teenage girls. These are children having children, who for nine months have fed their unborn infant Coke and potato chips, and who learn about the birth process in between the labor pains. It's really shocking that young people in America today grow up knowing more about the "walkmans" and about the carburetors of a car than they do about the characteristics of their own bodies.

If I were reshaping the school curriculum for century 21 to help students see connections, I'd have one major strand of study called the life cycle at the very core of common learning with the focus on nutrition, and on health and wellness, and with every student having an applied project, which would include the caring for of some form of life. I'm suggesting that being truly educated means learning about how one's own body functions, it means observing a variety of life forms. And above all, it means reflecting sensitively on the mystery of birth and growth and death. Without such knowledge, we remain ignorant and violence invades our culture on every side. This brings me to priority number two.

## Symbols and Their Uses

In addition to the life cycle, all people on the planet also use symbols to express feelings and ideas. After our first birth, we utter sounds and start reaching out to one another. Language comes immediately after life itself. In fact, Kay's convinced me that language begins *in utero*, as the unborn infant monitors the mother's voice. I think it's no accident that the three middle ear bones—the hammer, the anvil, and the stirrup—are the only bones that are fully formed at birth. We start listening before we speak. Quality education surely means becoming proficient in the use of symbols, in the written word, in the spoken word, and in discovering the numeracy, which is a symbol system, too. The point is that our sophisticated use of language sets human beings apart from all other forms of life, and it's through the miracle of words that we're all connected— connected to each other.

Consider the miracle of this moment. I stand here vibrating my vocal folds, molecules are bombarded in your direction, they hit your tympanic membrane, signals go scurrying up your eighth cranial nerve, and there's a response deep in your cerebrum, which approximates, I trust, the images in mine. But do you realize the audacity of this act—something we just take for granted? I must say that I'm encouraged that you're looking in my direction, but I also must say that I've been a teacher far too long to confuse visual contact with cerebral interaction.

I'm suggesting that language is not just another subject. It's the means by which all other subjects are pursued; and the new curriculum should include a second strand called the use of symbols, which might include the history of language, a study of mathematics as a symbol system, and surely it would include speaking, and listening, and reading, and writing across the whole curriculum, since it's through clear writing that clear thinking can be taught.

Above all, I do believe that students in our schools should be asked to consider the ethics of communication, since good language means not just accuracy, it means honesty as well. If you cannot trust and have some reliance on the message that I send, then communication is disrupted and the use of words becomes a violent and destructive act. Today's students live in a world where obscenities abound. They live in a world where cliches have become substitutes for reason. Students in the nation's schools urgently need to be taught how to distinguish between deceit and authentic communication.

Just a year or so ago, I read an article written by Dr. Elton Trueblood, who for years was president of Earlham College and now is in his nineties, but his mind is still going 150 miles an hour. He wrote about his early days as a graduate student at Johns Hopkins. He said he had a professor who, in the Oxbridge tradition, used to assign a paper every week. Now that's heavy sledding. Then he'd critique that paper every week. Dr. Trueblood said he'd get the paper back with all the red marks that would criticize the footnotes and comment on the syntax, but always, he said, in the margin and underlined were the words, "Is it true?" "Is what you've written really true?"

We're often struggling with the spelling, and the syntax, and the structure, but at heart communication is authentic only when it can be relied on to carry honesty and integrity from the sender. I'm suggesting that to be an educated person means helping students understand that language is a sacred trust and that truth is the obligation we assume when we're empowered with the use of words. This brings me to core commonality number three.

### Aesthetic Responsiveness

Beyond the life cycle and beyond the use of symbols, all people on the planet respond to the aesthetic. Dance is a universal language. Architecture is a universal language. Music is a universal language. Painting and sculpture are languages that can be understood all around the world. I'd like to see a third strand devoted to responsiveness to the aesthetic. Isn't it amazing that Salvador Dali's painting *The Persistence of Memory* can profoundly communicate to any person haunted by the relentless passage of time? Isn't it remarkable how the gospel song "Amazing Grace" can stir a common bond among people whether from Appalachia or Manhattan? Isn't it inspiring how "We Shall Overcome," when sung in slow solemn cadence, can stir powerful feelings regardless of race or economic status?

I find it intriguing that when archaeologists wish to study the status of past civilizations, they examine the artifacts of art. They look at pottery. They look at cave paintings. They look at musical instruments to determine the quality of the culture. I wonder what future archaeologists will think of us as they dig up plastic and Coke bottles. The arts are, in my opinion, above all, a special language of the children. Even before they learn to speak, little children 2 and 3 years old or earlier respond intuitively to dance and to music, to color and to rhythm.

The arts are uniquely helpful, in my opinion, for children who are disadvantaged. For several years I taught children who were deaf. These children couldn't speak because they couldn't hear. They were hugely frustrated because there is something in us that wants to get the message out. We need connections. Deaf children could not sustain that social interaction, so they had fear and frustration, which they expressed in rage. But I often observed that when you'd put paint brushes in their hands, or when you'd provide music for them to feel, if not hear, they discovered the satisfaction of self-expression. I have often been inspired by the Gallaudet Dancers in this city who may not hear the sound, but feel the rhythm of the dance floor and themselves become magical in the messages that they send.

The simple fact is that every student who enrolls in school has the innate urge and, I'm convinced, the innate capacity to be artistically expressive. It's really tragic that for most children the universal language of the arts is suppressed and then destroyed in the early years of learning because school boards consider arts a frill. I'm suggesting that for the most intimate and most profound and most moving experiences in our lives we turn to music and to dance and to the visual arts to express feelings and ideas that words cannot convey. The arts are, as one poet put it, the language of the angels. To be a truly educated person surely means being sensitive and responsive to the universal language we call art, which should be a central strand of the curriculum in every school. This brings me to human commonality number four.

### Our Position in Time and Space

While we're all non-uniform, as the Stanford students put it, and while we differ dramatically from each other, the simple truth is that all people on the planet have the marvelous capacity to place themselves in time and space, which no other creatures on the planet have the capacity to do. We explore our sense of space through geography and astronomy. We explore our sense of time through history, and yet how often we squander this truly awesome capacity to look in both directions, even neglecting our own roots.

Looking back several years ago, it occurred to me that the most important mentor in my own life was my grandfather Boyer who, incidentally, lived to be 100 years old. When he was in his early nineties, I asked Grandpa one day what he thought of man walking on the moon. He looked at me with a twinkle in his eye and said, "Ernest, the moon was made to look at not walk on." I'm convinced that to his dying day, he thought moon walking was the fantasy of an overly imaginative grandson.

The point is, though, that Grandpa, at the age of 40, moved his little family into the slums of Dayton, Ohio, which as you know, is the cultural center of the free world—the place where I was born. He then spent the next 40 years running a city mission, working for the poor, giving to the needy, teaching more by example than by word that to be truly human one must serve—a lesson I didn't learn in school. I learned by looking at a grandpa who knew how to live a life, and yet far too many children do not have intergenerational connections.

Margaret Mead said on one occasion that the health of any culture is sustained when three generations are vitally interacting with each other—a kind of vertical connection in which the different ages are connected. Yet in America today, we're creating what might be called horizontal culture. With each generation living all alone, disconnected from the others. The infants are in nurseries, toddlers are in daycare, older children are in schools, and we layer them by age. College students are isolated on campuses, living in a climate of low-grade decadence. Adults are, at best, in the workplace, and older citizens are in retirement villages living and dying all alone.

For several years my own parents lived in a retirement village where the average age was 80. I called my father one day and he said, no big deal being 80 around this place. You have to be 90 just to get a cake—waiting for that big birthday party in the sky. The nice feature is that they had a daycare center there so that every morning 50 little 3- and 4-year-olds came trucking in. Every little daycare child had an adopted grandparent. They met with them every day and had a sense of bonding. When I called my father, he wouldn't talk about his aches and pains; he'd talk about his little friend, who he was sure was going to be governor or perhaps even president some day. When I'd visit him, he'd have the paintings and the drawings taped on the wall. There was something powerful about the authenticity across the generations, the connectedness of things. An older person being inspired by the energy and innocence of youth, and a little boy being inspired and learning lessons about the courage and agony of aging.

I am convinced, to return to my central theme, the time has come to try to break up the age ghettos of this culture. I don't think we will abolish the institutions that now have become isolated. I think we need to think about building intergenerational institutions, bring the old and young together in daycare centers, and even with grandteachers in the schools. In the new core curriculum, with a strand called time and space, we not only would have children look at themselves and orient themselves geographically, they should orient themselves in time as well, discovering their own roots, completing perhaps an oral history, and beyond their own extended families becoming well informed about the culture that surrounds them and learning about other cultures, too.

To put it as simply as I can, students should study Western Civilization to understand our past. They should study non-Western cultures to understand our future. But in a larger sense, I'm suggesting that a truly educated person is one who sees connections, one who can place herself or himself in time and space. This brings me then to human commonality number five, in case you're keeping score.

### The Social Web: Group Membership

In addition to the life cycle, and the use of symbols and the arts, and our shared sense of time and space, I'm convinced that all people on the planet also hold membership in groups and institutions that consequentially shape their lives. To be truly educated means learning about the social web—learning about the social web of our existence. We're born in institutions, we live in institutions, and we're buried by institutions. It means knowing something about family life, about governments and how they function, about the informal structures that surround us, and, above all, it means knowing how group life varies from one culture to another.

We have a son who lives in a Mayan village in the jungles of Belize with his Mayan wife and four children—the brightest and the most handsome grandchildren on the planet, and I have pictures, as you might suspect, to prove it. When we visit Craig each year, I am impressed and re-impressed that Mayans and Americans carry on their work in very similar basic ways. I've been fascinated to do a cross-cultural study of the institutions, just as an observer. The jungles of Manhattan and the jungles of Belize are separated by 1,000 miles and 1,000 years; and yet the Mayans, just like us, have their family units. They have elected leaders. They have village councils. They have law enforcement officers and jails, and schools, and places where they worship. At one level it is all very different, but at another level it is very much the same.

I'm suggesting that we all hold membership in groups and institutions, which could be a fifth strand in the new curriculum. I believe all students should not only be introduced to the web of institutions in their own lives, but also engage in a cross-cultural study, which might, for example, compare Santa Cruz, California, with Santa Cruz, Belize. This brings me to commonality number six.

### Producing and Consuming

The simple truth is that with all of our differences, all people on the planet spend their time producing and consuming—a sixth integrative theme in the curriculum I imagine. Through such an effort, students would understand the nature of work and perhaps again to be prepared for work itself.

The sad truth is that today's young people grow up in a culture preoccupied with consuming, with little understanding of what it means to actually produce. I give you as Exhibit A, Toys Are Us, which for an older man is a bewildering, and dazzling, and enormously frustrating experience with boxes of junk, and plastic, and dazzling things piled from floor to ceiling, as if they'd been poured out of some cornucopia of abundance. We know all about production, but where does it all come from? Today's students even see their parents bringing paper home at night and carrying more paper off in the morning, which seems, incidentally, to incestuously procreate overnight. The question children surely must be asking is, what is it that grownups actually do—besides go to airplanes and airports and the like?

When I was chancellor of the State University of New York, I took our youngest son, who was then eight, with me to our cabin in the Birches for a weekend. My goal was to build a dock. All day, instead of playing, Steven sat at the water's edge watching me do things he'd never seen me do before and hearing me say things he'd never heard me say before. That evening as we drove home, Steven remained quite pensive. Finally, after several miles, he said, "Daddy, I wish you'd have grown up to be a carpenter instead of you-know-what." He had no idea what you-know-what involved, but somehow he was enormously intrigued, and in some ways reassured, to see his father even crudely engage in production. That is, demonstrating that one can truly, in a craftsmanship sense, try to make something work.

I am convinced that a new integrated curriculum for the future might include a strand called producing and consuming, with each student studying simple economics and examining how money systems differ from one culture to another, learning how work differs from one society to the next, and completing a work project of their own to gain a respect for the life of craftsmanship itself.

Several years ago, when Kay and I were in China, we were intrigued that in many of the elementary schools they had a little coats factory just down the hall where children would go in the afternoon and make a product that would be sold in the local schools. We also were told about a student who had defaced the surface of his desk, and as a punishment, he spent three days in the factory where the desks were being made, helping the workmen and observing the painstaking effort that was involved. We were told, not surprisingly, that the student never defaced a desk again. I'm suggesting that the new curriculum might include a study of producing, consuming, and conserving—preparing students not just for college, but also for the exciting world of work. This brings me to commonality number seven.

### Our Connectedness with Nature

It's true that all people on the planet are different, but it's also true that we're all connected to the ecology of this planet in which we are, as Lewis Thomas put it, embedded as working parts. To be truly educated for the next century includes understanding our connectedness to nature. David, my 4-year-old grandson in Belize, understands these connections very well. He chases birds. He bathes in the river. He watches corn being picked, and pounded into tortillas, and heated over the open fire. But I must tell you that David's cousins, through no fault of their own, who live in Boston and in Princeton with the appliances, and with asphalt roadways, and precooked food, find it enormously more difficult to discover their connectedness to nature.

When I was United States Commissioner of Education, Joan Cooney, the creator of "Sesame Street," came to see me and said the Children's Television Workshop wanted to create a new program on science and technology for junior high school students so they could understand more about their connectedness to the natural world. It subsequently was funded and called "Three-Two-One Contact." You may have seen it. In doing background work for that project, Children's Television Workshop surveyed some junior high school students in New York City; and they were asked such questions as, where does water come from? A sizable percentage said, the faucet. They were asked where does light come from, and they said, the switch. When they were asked where does garbage go, they said, down the chute, which is better than out the window, but you get the point. I think it's not unfair to say that the sense of connectedness for most of our students goes about as far as the VCR, and the refrigerator door, and the light switch on the wall, which is a dangerous and ignorant relationship to the ecology of our planet.

I'm suggesting that with all of our differences, every single one of us is inextricably connected to the natural world. All students, during their days of formal learning, should explore this commonality by studying the principles of science, by discovering how technology has profoundly shaped their own lives, and, above all, by learning that our very survival on this planet means not only understanding, but preserving the earth home we share together.

### Values and Beliefs

Finally, all people on the planet, regardless of their unique heritage or tradition, are, I am convinced, searching for a larger purpose. We all seek to give special meaning to our lives. We cannot be, in an ethical, and moral, and spiritual sense, disconnected. Reinhold Niebuhr put it precisely when he said man and woman cannot behold unless they be committed. They cannot find themselves unless they find a purpose beyond themselves. When all is said and done, the truly educated person is one who establishes values and beliefs; or to state it in an old-fashioned way, as my Grandfather might put it, someone who has convictions.

During our study of the American high school, I became convinced that we have not just a school problem, but a youth problem, in this country. I was struck that far too many teenagers feel unwanted, and unneeded, and unconnected to the larger world; and without guidance and direction, they lose their sense of purpose at a very early age. Vachel Lindsay wrote on one occasion:

> It is the world's one crime that babes grow dull. Not that they sow, but that they seldom reap. Not that they serve, but have no God to serve. Not that they die, but that they die like sheep. The tragedy is not death. The tragedy is to die with commitments undefined, with convictions undeclared, and with service unfulfilled.

With all of the controversy that surrounds it, I remain convinced that a school is a place where values are examined, not that answers are dictated; but that it provides a climate that makes honorable the quest. I also am convinced, as I look back on my own schooling, that the search for meaning was not taught so much in the curriculum, but it seemed to be taught most effectively by great teachers who modeled values in their own lives.

Several years ago I couldn't sleep, and instead of counting sheep, I counted all the teachers that I'd had. There were a few nightmares in the bunch, but I can tell you that there were three or four or more I remembered with great reverence because they profoundly changed my life. I remember a university professor who in a Shakespeare seminar used to read Shakespeare aloud in class, *King Lear* and *Macbeth*, teaching me that literature is not an ancient language, it's an inquiry into the deepest yearnings of the human spirit. I remember Mr. Whitlinger, a high school teacher, who said one day in class, Ernest, you're doing very well in history. Keep this up, and you just might be a student. It was the highest academic accolade I'd ever had. You mean, I'm not a cowboy? I'm something Mr. Whitlinger calls a student? I remember my 1st grade teacher, who the first day of school said, "Good morning, class. Today we learn to read." Fifty years later, I wrote a book on high school. I had a chapter called the centrality of language, and it occurred to me that it wasn't just an accident. It was the influence of an unheralded 1st grade teacher, Fairview Avenue Elementary School, whose influence lives on forever, like ripples in a pond.

What made these teachers truly great? I asked myself one day. It occurred to me there were four characteristics that they all shared in common, even though they were very, very different.

Every teacher I had was knowledgeable and well informed. There was something there to teach, but that doesn't make a great teacher or one whose influence is enduring. I've had teachers who were knowledgeable, but were terrible in their teaching. The second characteristic was that all teachers were not only well informed, they related what they knew to the readiness of students. Third, all the great teachers that I had created active, not passive, learning in the classrooms. They occasionally became learners while the students became the teacher, but that still didn't make these teachers truly great. It occurred to me that the most influential teachers that I had were the ones who were authentic, and honest, and open human beings. They taught not only their subjects, they taught themselves. They were willing to be vulnerable. They laughed, they cried, they said, "I don't know." As I observed them, I was observing not just the content to be covered, but I observed the quality of their lives.

Values are taught, in my opinion, most especially by the integrity and the honesty, and the openness, and the authenticity of the teacher. But I think values also can be taught through service, and I do believe that all students should be asked to complete a community service project in order to demonstrate that learning has a practical and applied side that must be tested. I'd like to see all students complete a service project in daycare centers, or in retirement villages, or in tutoring other children in school since, obviously, one way to learn, and perhaps the best way, is to try to teach. Martin Luther King said, on one occasion, that everyone can be great because everyone can serve. I'm convinced that the young people of this country are ready to be inspired by a larger vision.

## A Vision for Our Children

But then—and now I come to the final question again—what then does it mean to be an educated person? It means respecting the miracle of life. It means being empowered in the use of language. It means responding aesthetically to the artistic. Being truly educated means understanding our membership in groups and institutions. It means having reverence for the natural world. It means affirming the dignity of work. And above all, being an educated person means being guided by values and beliefs and connecting the lessons of the classroom to the realities of life. These are the core competencies that I believe should replace the old Carnegie units. All of this can be accomplished as the schools focus, not on seat time, but on students in classrooms that are true communities of learning.

I know how idealistic it may sound, but it is my urgent hope that in the century ahead, students in the nation's schools will be judged, not by their performance on a single test, but by the quality of their lives and by the hopefulness of their potential. It is my hope that students in the classrooms of tomorrow will be encouraged to be creative, not conforming, and learn to cooperate rather than compete. It is my abiding hope that students in our schools will begin to see the world as whole and be inspired by the beauty, as well as the challenges, of the world around them.

Above all, it is my urgent prayer that Julie, my 5-year-old granddaughter who lives in Princeton, and David, her 5-year-old cousin in Belize, along with all of the other children on this planet, will grow up knowing down deep inside that they are truly members of the same human family to which we are all inextricably connected. Fifty years ago, Mark Van-Doren wrote that the connectedness of things is what the educator contemplates to the limit of his capacity. VanDoren concluded by saying that the student who can begin early in life to see things as connected has begun the life of learning. This, it seems to me, is what being an educated person is all about.

I noticed in your program that the theme of this Conference is not only communities of learning, but a commitment to the world's children. In conclusion, I thought I might mention a prayer that was sent to me by Marion Wright Edelman that seems to fit, with some moderation, the Conference theme that launches this convention. The prayer reads:

> Dear Lord, we pray for children who cannot find their shoes. We also pray for those who have no shoes. We pray for children who squirm in church and synagogue, who eat dessert before dinner, and who never pick up their rooms. We also pray for children who rarely see a doctor, who never see a dentist, and who have no rooms to pick up, and whose pictures are not on anybody's dresser. Dear Lord, we pray for children who are always loved, and we pray especially for children who will grasp the hand of anyone kind enough to hold them.

I suggest, then, that our obligation in the century ahead is to commit ourselves not just to build better schools, not just to build a better world, but above all, to build a better climate for our children.

# IV

# *Aligning Assessment with Curriculum and Instruction*

# IV-1

# Aligning Assessment with Curriculum and Instruction

## Introduction

Judith A. Arter
Northwest Regional Educational Laboratory

### Alignment

Alignment is a simple and practical idea. If we want to maximize the impact of instruction, we must have a clear idea of what we want to accomplish and focus instruction on attaining those goals. Then, assessments should provide a clear understanding of how we—teachers and students—are doing on attaining those goals. Everything from instruction to curriculum and assessment reaches toward the same goals. **Costa** and **Kallick** state it this way:

> Aligning the curriculum is based on three major decisions: (1) establishing the purposes, outcomes, goals or objectives of the educational enterprise . . . ; (2) designing the delivery system by which those goals will be achieved, including instructional design, materials selection, allocation of time and placement of learning; and (3) developing feedback loops for monitoring, collecting evidence of, and evaluating the achievement of our goals as a result of employing that delivery system.

The notion of alignment is as valid at the classroom level as at the school, district, or state level. Fielding and Schalock (1983) cite the steps involved in alignment in the classroom: (1) identify the learning outcomes students are expected to achieve; (2) select or develop instructional materials and procedures likely to foster outcome attainment; (3) make clear to students the nature of the learning outcomes they will be working toward and on which they will be assessed; (4) keep track of progress toward the outcomes during instruction; (5) provide feedback to students on progress; and (6) use assessment information to improve instruction.

The general concept of alignment is sound; however, many practical questions arise when attempting the alignment process:

1. How does the notion of alignment at the school, district, or state level translate into teachers' day-to-day activities?

2. What is the best way to ensure that student learning improves through explicit feedback loops?

3. What does "integrating assessment and instruction" mean? The special interest group on classroom assessment of the American Educational Research Association polled members in an informal survey to "determine the most important issues and problems confronting classroom assessment and its practitioners." The most frequently cited topic was "helping teachers link assessment to day-to-day instruction" (Airasian 1993). Yet the meaning of the phrase "linking (integrating) assessment and instruction" is not always clear.

4. How do large-scale assessment (done outside the classroom) and classroom assessment interrelate? How does alignment outside the classroom relate to alignment within the classroom?

5. Which comes first after goals are established—curriculum and instruction, or assessment? How does assessment influence curriculum and instruction? Does assessment benignly reflect instruction and record progress toward goals, or can assessment "push" curriculum and instruction in desired directions?

Authors of the papers in this section provide insights on these issues. For example, the complexities and subtleties of these questions is well demonstrated in the paper by **Dennie Palmer Wolf, Paul G. LeMahieu and JoAnne Eresh**, who give examples of necessary changes in teachers' day-to-day activities if we are to realize for students the high standards set in national policy. They say

> "God is in the details." Ultimately, the realization of these national policies on standards and assessment will be played out not in Washington but in the way a teacher answers a child's question; the way a principal finds joint planning time for teachers; and the way a community of parents and teachers joins forces to see that these standards become a tool for, rather than an obstacle to, equity.

The message is that large-scale and classroom activities *do* relate, often in complex and subtle ways. Also, if we want to achieve real reform, we have to assist teachers in reexamining their day-to-day activities.

## Integrating Assessment and Instruction

The papers in this section represent four different views on how assessments fit into the alignment process, and can also be seen as general views of what it means to "integrate assessment and instruction."

### Assessment as Continuous Monitoring

The first view of integrating assessment and instruction is that assessment provides the information to power instructional decision making. It is the part of the feedback loop that helps us to monitor student progress in order to make adjustments to instruction. At the large-scale level we monitor student progress to make decisions about the quality of education. At the classroom level teachers need information about the status of student progress toward goals to effectively plan instruction.

Using assessment to monitor student progress is not new. What might be new are: (1) the types of student outcomes on which we monitor progress (with more emphasis on process skills and the outcomes of collaboration, critical thinking, taking control of learning, and so on); (2) the need for more performance-based assessments to monitor progress on these outcomes; and (3) the need to collect more systematic information in the classroom.

Several papers in this section relate to assessment as monitoring. **Jackie Palmer** describes new thinking about science and mathematics instruction and provides examples of how alternative-assessment information fits into the instructional cycle in the classroom. **Fitzpatrick** and her colleagues describe a districtwide cross-curriculum, computer-based system that makes continuous monitoring more efficient in the classroom. **Christine Sobray Evans** provides a personal view of monitoring in her own classroom.

### Assessment as a Tool to Push Instruction in a Desired Direction

The second view of integrating assessment and instruction is based on the position that large-scale assessment communicates to teachers what and how to teach. Therefore, if we design our large-scale assessments correctly, we will drive instruction in the desired direction. Gong, Venezky, and Mioduser (1992) describe a study they did in science classrooms in which performance-assessment tasks were explicitly designed to model current theories of good instruction—long-term projects, clear outcomes, emphasis on the process of science, student collaboration, and so on. The authors argued that this type of modeling is necessary because many teachers don't know how to teach science to attain the types of goals currently proposed for students. They found an impact on teachers in terms of their: (1) perception of science, (2) knowledge of science, (3) awareness of students, (4) teaching styles and methods, and (5) professional interaction.

The **Richard J. Shavelson and Gail P. Baxter** paper, based on assessment in science and mathematics, provides another good statement of this view. They propose that current knowledge about how students learn and the changing nature of our goals for students require different instructional techniques, classroom configuration, and assessment. Because teachers "teach to the test," if we design certain features into our assessments, teachers will have to alter their instruction if only to ensure their students do well on the test. The paper specifically discusses current knowledge about how students learn, the changing nature of our goals for students in science and math, and, as a result, features that need to be designed into assessment tasks.

Building certain features into our large-scale assessment tasks may not be enough, however, to change classroom instruction. Based on experiences in Kentucky, **Thomas R. Guskey** relates that "teachers did not know how to teach to the performance-based assessments, nor did they believe that they could do so within their current time constraints." One way to maximize the impact of our performance-assessment systems, then, is to provide additional staff development on how to model performance tasks in the classroom.

The second view of integrating assessment and instruction thus emphasizes the impact of assessment *tasks* on instruction. Teachers model subsequent instructional activities after the *activities/tasks* in the assessment. For example, once teachers see how to incorporate collaboration into an activity, they apply it to later classroom activities.

### Assessment *Development* as a Strategy to Change Instruction

A third view of integrating assessment and instruction is similar to the second view except that it is the *development* of the assessment (in addition to the *use* of the assessment) that changes instructional practice. When writing about their experiences in developing meaningful alternative assessments, author after author says that the process of developing the assessment is as useful as having the assessment in place at the end. Statements such as the following are typical:

> The desired outcome is fully realized only when the participants themselves work through the issues critical to achieving it. The processes of exploration and search are important components in changing educational practice; it is not enough to pass on the wisdom of others' experience (*Portfolio* 1990).

> The dialogue that is happening among teachers is one of the most exciting things happening from this. Portfolios may actually bring us together more than ever before so we begin to define our criteria much more precisely (Sugarman 1989).

> The benefits of portfolios lie as much in the discussions they generate among teachers—and between teachers and students—as in the wealth of information they provide (Murphy and Smith 1990).

Or, as **Evans** states in this section,

> To our surprise, the [model student papers we chose to illustrate points on the scoring guide] were just one benefit of this entire process. More important, we felt, were the discussion and negotiation that occurred that afternoon. In addition to learning about student writing, we valued the opportunity to talk to one another about the work we do. That sense of being a community of learners continued for the entire year.

To develop a meaningful alternative assessment, one needs expertise in the content to be assessed, in how learning occurs, and in assessment. The person who develops the assessment does the learning. Why not make teachers these developers so they can take this increase in expertise back to the classroom? **Costa and Kallick** are clear on this point. If we want meaningful change in assessment, we need to reenfranchise teachers and students by giving them the power to assess progress; by supporting schools in becoming learning organizations; by putting teachers back in control of developing assessments; and by developing an "intellectual ecology" of the school environment. The development of the performance *criteria* rather than the performance *tasks* is frequently cited as having this positive impact on teaching expertise.

### Assessment as a Tool for Learning

A fourth view of integrating assessment and instruction is based on the belief that both large-scale and classroom assessment influence students directly—for better or worse. Assessment is not just a neutral episode of collecting information (as in the first view), nor just a way of influencing future instruction by teachers (as in the second and third views), but is a way of directly influencing students—in positive ways if designed properly. For example, students could learn something from doing the assessment—the activity itself is an instructional episode.

One view of how this influence works is presented in the paper by **Francine Z. Stayter and Peter Johnston** in which they detail the many ways students get messages from our assessments—what to study, who has control of judgment, students' role in the classroom—and specific examples of how to design assessments that send productive messages. Although their examples come from the area of literacy assessment, their comments apply to other areas as well.

**Costa and Kallick** present a similar theme, stating, "We must constantly remind ourselves that the ultimate purpose of evaluation is to have students become self-evaluative."

**Judith Arter** articulates another slant on the influence of assessment on students. She uses examples in writing, reading, and science to show how well-articulated performance criteria and blueprints for paper-and-pencil tests can promote student learning. Arter states that assessments are learning episodes when they help teachers and students clarify the expected outcome of instruction and promote student self-evaluation and reflection. In this view, performance assessments influence students through the use of *performance criteria*. **Doris Sperling** makes the same point when she describes one experience in which a teacher assisted her students in articulating and using criteria for writing.

A third way that assessments can directly influence students is illustrated by three papers in this section. (In a fourth paper not included, Gong, Venezky, and Mioduser (1992) describe how the *tasks* used in their assessment were instructional—students learned from the extended projects that were the basis for the assessment.) **Roberta Camp**'s paper is an excellent example of the instructional power of portfolios, which she sees as more than a receptacle for evidence of student progress. When approached properly, the process of assembling portfolios includes occasions for self-reflection and self-evaluation, and it helps students develop such student skills as critical thinking, metacognition, planning, and taking control of learning.

The paper by **Jane B. Braunger** describes how students' retelling of stories is a good way to assess student comprehension *and* give students practice in both determining important points to retell and summarizing them for the listener. Again, the retelling *task* that students perform for assessment is also an episode of learning. The **Evans** paper has a similar theme.

Thus, alternative assessments can have a direct impact on students through the messages they send, through appropriate use of performance criteria, and from completing the assessment task itself.

## Implications for Assessment Design and Development

Considering the various views of "alignment" and "integrating assessment with instruction" is interesting and informative, but it is also imperative. One's philosophy and unstated assumptions about "integration" will affect how alternative assessments are designed and developed. For example, if *you*, the large-scale assessment developer, would like *your* assessment to push instruction in a new direction, then you must enlist teachers in the development and you must have a clear idea of necessary instructional changes so they can be designed into the assessment.

Likewise, if *you*, the teacher or large-scale assessment developer, want to use assessment as a direct tool for student learning, *you* must design certain features into the assessment. For example, **Arter** proposes that to increase the instructional usefulness of performance assessments, performance criteria should be designed as generalizable across tasks—to judge, for example, many student essays and not just the current essay. If we design our scoring procedures to be task-specific, students will not be as likely to generalize them to the next task. If our goal is to create students who "can win again," what good does it do to create mechanisms for them to see how to "win" on only the current task. (See the **Arter** papers in other parts of this book for additional discussion of this topic.)

## Conclusion

The 12 papers in this section are similar to one another in that they (1) view alternative assessments as necessary tools for aligning curriculum and instruction with many of the goals we have for students; (2) view alternative assessment as aligning well with current theories of how students learn; and (3) propose a more active role for teachers in the development and administration of large-scale assessment. Although they differ in their vision of how the alignment process plays out in terms of integrating assessment and instruction, these views are not necessarily mutually exclusive.

### REFERENCES

Airasian, P.W. (December 1993). "Issues and Problems in Classroom Assessment." *Classroom Assessment Newsletter* 4, 2.

Fielding, G., and D. Schalock. (1983). *Integrating Teaching and Testing: A Handbook for High School Teachers.* Eugene, Oreg.: Teaching Research Division, Oregon State System of Higher Education.

Gong, B., R. Venezky, and D. Mioduser. (September 1992). "Instructional Assessments: Lever for Systemic Change in Science Education Classrooms." *Journal of Science Education and Technology* 1, 3: 157-176.

Murphy, S., and M. A. Smith. (Summer 1990). "Talking About Portfolios." *The Quarterly of the National Writing Project* 12, 2: 1-3.

*Portfolio—The Newsletter of Arts PROPEL.* (November 1990). 1, 1.

Sugarman, J. (November 27, 1989). "Portfolios Useful Assessment Tool." *Education USA* 32, 101.

# IV-2

# Reassessing Assessment: Seven Issues Facing Renaissance Schools

Arthur L. Costa
California State University, Sacramento

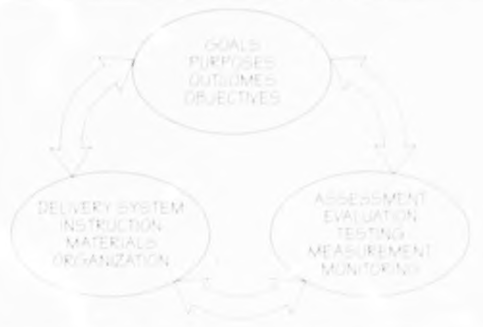
Bena Kallick
Westport, Connecticut

## The Nature of Alignment

Aligning the curriculum is based on three major decisions: (1) establishing the purposes, outcomes, goals, or objectives of the educational enterprise at the classroom, school, district, state, and national levels; (2) designing the delivery system by which those goals will be achieved, including instructional design, materials selection, allocation of time, and placement of learnings; and (3) developing feedback loops for monitoring, collecting evidence of, and evaluating the achievement of our goals as a result of employing that delivery system.

These three decisions are made at a variety of levels. For example, setting student learning goals occurs at these levels

FIGURE 1
CURRICULUM ALIGNMENT


GOALS
PURPOSES
OUTCOMES
OBJECTIVES
DELIVERY SYSTEM
INSTRUCTION
MATERIALS
ORGANIZATION
ASSESSMENT
EVALUATION
TESTING
MEASUREMENT
MONITORING

- At the national level: In 1990, then-President Bush produced a document establishing six goals for schools for the year 2000 based upon a conference of state governors.

---

Adapted from *If Minds Matter: A Foreword to the Future, Vol. II*, edited by Arthur L. Costa, James Bellanca, and Robin Fogarty. Copyright © 1992 by IRI/Skylight Publishing, Inc., 200 E. Wood Street, Suite 274, Palatine, Illinois. Reprinted with permission.

BEST COPY AVAILABLE

- At the state level: The state adopts goals, performances, and standards for students at third, sixth, eighth, and eleventh grades.
- At the district level: School district trustees adopt a mission statement and student outcomes. They may adopt lists of performances and graduation requirements.
- At the classroom level: Teachers decide what should be grade-level learning outcomes in reading and math for fifth grade students. The teacher plans objectives for the science unit.

Likewise, the delivery system is designed to achieve those goals

- At the national level: Talk of a national curriculum, Head Start, Chapter 1, etc.
- At the state level: Numerous states develop curriculum guides and adopt textbooks, mandate the number of required hours of instruction, etc.
- At the district level: Decisions are made about staff development, allocation of resources, school building design, evaluation of instruction, etc.
- At the classroom level: Teachers devise instructional strategies, time sequences, and classroom organizational structures intended to facilitate achievement of their goals.

Similarly, procedures for assessing the attainment of student learning goals are developed:

- At the national level: The National Assessment of Educational Progress discloses how U.S. students score on a variety of topics.
- At the state or district level: States and districts adopt large-scale testing procedures to assess students' achievement of the state or district goals.
- At the classroom level: Teachers devise tests and other forms of observation of student performance of desired behaviors.

## Learning Goals for Students

In the curriculum alignment process, sound educational practice dictates that the first group of decisions (the goals) needs to drive the system. Like it or not, what is inspected is what is expected; what you test is what you get.

Educators realize that new goals for the next century are becoming increasingly important as survival skills for our children's future, for the continuity of our democratic institutions, and even for our planetary existence. Such goals include

- Capacity for continued learning
- Knowledge of how to behave when answers to problems are not immediately apparent
- Cooperation and team building
- Precise communication in a variety of modes
- Appreciation for disparate value systems
- Problem solving that requires creativity and ingenuity
- Ability to resolve ambiguous, discrepant, and paradoxical situations
- Generation and organization of an overabundance of technologically produced information
- Pride and craftsmanship of product
- High self-esteem
- Personal commitment to larger organizational and global values.

These new goals need to drive curriculum alignment in thoughtful schools of the future. The delivery system—curriculum materials, instructional strategies, school organization, and the curriculum alignment decision-making processes employed—needs to embody these goals not only for students but for all of the schools' inhabitants.

## Rationale for Alternative Assessment

Likewise, our methods of assessment must be made consistent with our new goals. We cannot employ product-oriented assessment techniques to assess the achievement of these new, process-oriented educational outcomes. Norm-referenced standardized test scores give us a static number that reflects the achievement and performance of isolated skills at a particular moment in time. Thinking, however, is dynamic: we learn from experience, react emotionally to situations, experience power in problem solving, and are energized by the act of discovery. Thus, testing thinking may indeed be an oxymoron.

We are witnessing a nationwide surge to "go beyond the bubble" of traditional standardized, norm-reference, computerized testing (California State Department of Education Curriculum Assessment Alignment Conference 1989). State departments of education (New Jersey, Vermont, Colorado, Maryland, Michigan, and Illinois, for example) are providing leadership by experimenting with and advocating innovative assessment methods such as writing samples, manipulation of materials, open-ended multiple-answer questions, portfolios, performances, and exhibitions.

Such innovative methods are more useful and "authentic" than traditional testing procedures for several reasons. First, they resemble the situations in which real problem solving and creativity are demanded: They are not contrived. Second, they allow teachers to more accurately diagnose students' abilities. Third, they take place during instruction rather than after instruction and provide more immediate results that assist teaching teams in evaluating the effectiveness of their own curriculum decisions and instructional efforts. Finally, they provide "real-time" feedback to students themselves who are (or who must become) the ultimate evaluators of their own work and performance.

## Steps to Ensure Success in Restructuring

To ensure the success of new assessment efforts in the restructured school, educators need to address at least the following seven tasks.

### 1. Re-establish the School-site Team as Locus for Accountability

For too long, the process of assessment has been external to teachers' goal setting, curriculum and instructional decision making. School effectiveness, student achievement, and teachers' competence have often been determined by a narrow range of standardized student achievement test scores in a limited number of content areas, for example, reading, math, and language acquisition. Test results have been published in rank order in newspapers. Awards of excellence have been granted to schools that show the highest gains in scores. Teachers receive merit pay based on their students' performance on standardized tests.

Lord Kelvin is quoted as saying, "When you cannot measure it, when you cannot express it in numbers, your knowledge is a very meager and unsatisfactory kind." In the effort to be more "objective," we have tried to translate our educational goals into observable, measurable outcomes that do not depend on teacher judgment. We have become fascinated and enamored with such things as these:

- The amount of time on task
- Gain scores on achievement tests
- Class size, numbers of students, and ratio of students to adults
- Length of time in school
- Numbers of days in attendance
- Percentages of objectives attained
- School effectiveness based on published test scores.

In the process, teachers have been disenfranchised. They have had little say about what data are collected, including test content. In fact, the consensus that traditional tests do not measure all that is important on the curriculum, and the results of testing disclose incomplete information about the adequacy of the curriculum and instructional decisions is increasing. In many ways, the desire for "objective" outcomes has signaled teachers that they are "incompetent" to assess students themselves. In effect, they were told they could not be trusted to collect evidence of students' growth; their daily classroom observations were suspect and of little worth.

The accountability movement has caused educators to search for "hard data" by which to assess their efforts. What teachers observed, therefore, was considered "soft data." We propose that the "hardest," most objective data available are collected by an enlightened teaching team that systematically and collectively gathers data over time in the real-life, day-to-day interactions and problem solving of the classroom. Conversely, the "softest," most suspect data are those designed and collected by testing "experts" external to the school setting and ignorant of the school's mission, values, and goals; the community's culture and socioeconomics; and the classroom's mix of learning styles, teaching strategies, and group dynamics. Someday we will reflect on this era and think: "What was educationally significant but difficult to measure was replaced by what was insignificant but easy to measure." As Jacob Viner said, "When you can measure it, when you can express it in numbers, your knowledge is STILL of a meager and unsatisfactory kind!"

In the restructured school, staff members need training to learn how to design multiple ways of gathering such data, establish criteria for judgment, work together to develop a common understanding of student outcomes, ensure reliability of observation, and report results effectively. Accountability goes hand in hand with informed, reflective practice.

### 2. Develop New Identities and Roles

Becoming a learning organization means transforming the identities of the participants in the process: students, teachers, administrators, coaches, etc. Because learning organizations honor and strive for the intellectual growth for which all human beings are capable, because learning organizations honor human diversity which all living creatures are genetically constructed to express, and because learning organizations strive to empower others to become increasingly self-modifying, these principles will pervade the classroom, boardroom, and conference room.

Over time, students will find that their role is becoming one of "envisioners" of their own desired states—establishing goals, making plans, and clarifying outcomes for themselves. They will understand that they are in charge of developing their own strategies for achieving those goals; that they will generate ways of assessing their own growth toward those ends; that they will view their role as knowledge producers and meaning makers for themselves and others; and that they can contribute to a cooperative environment for others to grow toward their desired state.

Teachers will reconceptualize their role as mediators of students' learning. Teachers, working in collaborative teams, will envision those desired states for themselves and their students and for the classroom climate they share. Teachers will generate strategies for achieving those desired goals; they will view themselves as knowledge managers, knowledge producers, and meaning makers; and they will generate data as feedback to guide and assess their own, their students', and the classroom climate's progress toward these desired states.

Staff developers will envision their role as mediators of teachers, staff members, and the community's continued learning. They will generate strategies for assisting teachers to be in a continual learning state. They will view themselves as knowledge producers and meaning makers, and they will generate data as feedback to assess their own, their teachers', and the school environment's progress toward these desired states.

Administrators will envision their role as mediators of the district and communitywide conditions for continual learning. They will cooperatively generate strategies for achieving the vision of a learning organization, and they will generate data as a means of assessing progress toward that vision. Their role will be analogous to that of "Environmental Protection Agency"—constantly monitoring the intellectual ecology of the district, so that cooperation and intellectual growth become neither threatened, nor endangered, nor extinct.

The role of the district trustees will be that of community educators keeping and protecting the value system of the district and monitoring their decisions to be consistent with these desired states.

### 3. Expand Range, Variety, and Multiplicity of Assessment Techniques

For too long we have relied on a limited range of acceptable measures, primarily paper-and-pencil tests. Assessing the goals of the restructured school will require an expansion of our repertoire of assessment techniques.

- Direct observation of student performance in problem-solving situations
- Portfolios of selected student work over time
- Extended projects
- Logs or journals
- Interviewing students about their self-concepts and perceptions of themselves as learners
- Videotapes of student interactions
- Writing samples
- Performance assessment using a set of agreed-upon criteria
- Anecdotal records
- Critical incidents
- Employing technology to assist in collecting and recording information about students over time.

### 4. Systematizing Assessment Procedures

Actually, skillful teachers already assess their students in many of the above ways. Teachers can readily describe their intuitive feelings of students progress: the light that goes on in students' eyes, the voice inflections of students when they've "got it," the "aha" of discovery, the collaborative sharing and interdependence of group work, the serendipitous application of skills and concepts beyond the context in which they were learned, and the reports from parents and other teachers of learnings applied.

What is lacking is a systematic way of collecting and reporting such evidence. Staff members need to refine their skills of observation and work for agreement in judgments. They need to identify and define terms and adopt common goals. They need to scrutinize the curriculum continually to ensure that goals, instruction, and assessment are aligned. Teachers, parents, administrators, and students will all need to see the school's objectives and purposes clearly, and all will become more involved in collecting data, revising perceptions and realigning practices.

### 5. Reeducate Legislators, Parents, Board Members and the Community

As a goal of education, the development of the intellect is not yet valued or understood by the majority of the public. While business and industry leaders are becoming more supportive of the school's endeavor to educate the future generation to become better individual and group problem solvers, to work cooperatively as a team members, to develop creative capacities, and to be open to new and continued learning, legislators, governors, school boards, and even an "Education President" still fail to include thinking, creativity, and cooperation in our national goals for the 21st century.

Educators, in conjunction with test makers, textbook publishers, professional organizations, business and industry leaders, parents, and the media, must begin a massive information and educational program intended to shift public policy and national values toward support of a more rational, cooperative, and compassionate mission statement for public education.

### 6. Assessment of School and Classroom Conditions as Part of Student Performance Assessment

Fritz Perls is quoted as saying, "Abnormal behavior is normal behavior under abnormal conditions." The conditions in which people work send powerful signals as to the values, mission, and purposes of the institution. Jack Frymier (1987) writes:

> In the main, the bureaucratic structure of the workplace is more influential in determining what professionals do than are personal abilities, professional training, or previous experience. Therefore, change efforts should focus on the structure of the workplace, not on the teachers (p. 10).

Teachers are more likely to teach for thinking, creativity, and cooperation if they are in an intellectually stimulating, creative, and cooperative environment themselves. We must, therefore, also assess the quality of the environment in which teachers, students, and administrators operate. Some questions by which to assess the intellectual ecology of the school environment include these:

- Has the curriculum been developed and materials adopted to support thinking, cooperation, creativity, and the other new goals of the restructured school?
- Are the reward systems for students and the evaluation system for teachers aligned with the new values and mission of the restructuring school?
- Do the communications such as report cards, newsletters to parents, school mottoes, and logos reflect the values of the restructuring school?

### 7. Encourage Students' Routine Self-evaluation

We must constantly remind ourselves that the ultimate purpose of evaluation is to have students become self-evaluative. If students graduate from our schools still dependent upon others to tell them when they are adequate, good, or excellent, then we've missed the whole point of education. The highest level of Bloom's Taxonomy (1956) includes generating, holding, and applying a set of internal and external criteria. For too long, adults alone have been practicing those skills. We need to gradually shift that responsibility to learners. Our goals of the restructuring school must be to help students and teachers develop their self-renewing capacities: self-analysis, self-evaluation, and self-modification.

## REFERENCES

Bloom, B.S., M.D. Englehart, E. J. Furst, W.H. Hill, and D.R. Krathwol. (1956). *Taxonomy of Educational Objectives, Handbook I, Cognitive Domain.* New York: David McKay.

California State Department of Education. (1989). Curriculum Assessment Alignment Conference, October 16, 1989. Sacramento: Author.

Frymier, J. (September 1987). "Bureaucracy and the Neutering of Teachers." *Phi Delta Kappan* 69, 1: 9-14.

# IV-3

# The Instructional Resource Network: Supporting the Alignment of Curriculum, Instruction, and Assessment

Kathleen A. Fitzpatrick
National Study of School Evaluation

Marilynn Kulieke, John Hillary, Vern Begitschke
Township High School District 214, Arlington Heights, Illinois

Over the past eight years, Township High School District 214 has developed and implemented a comprehensive, outcome-based education instructional system that includes restructuring the district's curriculum in terms of the learning outcomes essential for students' success, aligning instruction and assessment with these learning outcomes, revising district policy to support the development and implementation of the outcome-based instructional system, and developing a variety of support systems.

Two key policies have been critical to the development of our outcome-based instructional system. First, the district's graduation requirements are now defined in terms of essential learning outcomes rather than a simple list of the number of credits students must accumulate. Under the revised graduation requirements, students must demonstrate each of the district's general learning outcomes (e.g., problem solving and communication skills) and essential outcomes in each discipline or program area (e.g., mathematics, language arts, science). Over the past several years the curriculum has been restructured to provide students with a variety of opportunities to achieve these essential learning outcomes. The second key district policy that has been developed and implemented pertains to the district's outcome-based instructional model (see Figure 1).

FIGURE 1

DISTRICT 214 INSTRUCTIONAL MODEL

1. District teachers will establish course outcomes to be the basis for decisions regarding what will be taught. Individual teachers will identify unit and lesson outcomes for students prior to each unit of instruction and each lesson.

2. Teachers will design instructional strategies that will maximize student achievement of the outcomes and include multiple learning opportunities for students. Teachers will provide both frequent feedback to students concerning their achievement and additional learning opportunities when necessary. Teachers will use a wide variety of instructional strategies based upon the nature of the outcomes and the students' learning styles and needs.

3. Teachers will reference their assessments of student achievement (tests, essays, projects, portfolios, performances, etc.) to course outcomes. Assessments can be drawn entirely or partially from a districtwide bank of teacher-developed and teacher-selected items referenced to the course outcomes.

## Overview of the Instructional Resource Network

The Instructional Resource Network (IRN) is a computer-based integrated network with four main components. First, IRN contains a curricular listing and mapping of each learning outcome students are expected to achieve. Second, a clearinghouse of instructional strategies and learning activities is aligned with the outcomes. Teachers from across the district have collaborated in developing learning outcomes and curriculum design—which led to the development of the first component of the network—and have also contributed their teaching ideas for a variety of learning activities they have found helpful in promoting student achievement of outcomes. This collection of teaching ideas and learning activities are listed according to their alignment with the outcomes and also in terms of the various learning styles they address.

The third IRN component is a variety of faculty-developed, performance-based items, strategies, and rubrics that measure students' achievement of the learning outcomes. The fourth component is the analysis component and designed to aid in the review and evaluation of assessment data that, in turn, can lead to the renewal of the instructional process within a data-driven decision-making framework. For instance, by analyzing data, educators can determine if a refinement in instructional strategy is appropriate or if students need to master certain prerequisites to be successful.

Each of the IRN's components is designed and developed to serve three major goals. The first goal has been to provide a working example of the district's outcome-based instructional model. Initially, the IRN provided tangible illustrations of the alignment of curriculum, instruction, and assessment to support the early phases of the development and implementation of the district's outcome-based instructional model. Currently, the IRN supports the ongoing development and expansion of assessments of student learning and refinement of learning outcomes. Furthermore, the IRN is an available resource to assist new teachers in implementing the district's instructional model.

A second IRN goal has been to address the workload issues encountered in developing and implementing a comprehensive outcome-based instructional system. Through the IRN, teachers can access ideas and strategies that other members of the faculty have found helpful, rather than teachers each carrying the full responsibility for designing their own instructional strategy, learning activity, or assessment.

The third goal is to break down some of the isolation that teachers experience, especially at the secondary school level. Through the IRN teachers have greater opportunities to collaborate in developing curriculum and in designing instructional and assessment strategies. Moreover, the IRN has provided a means to honor the contributions of successful strategies from teachers across the district. To help clarify the IRN components, the following sections provide a description of the outcome-based curriculum, instructional clearinghouse, assessment bank, and analysis functions. In the closing section we offer our reflections on the current stage of development of the IRN steps, together with other elements under consideration.

## The Outcome-Based Curriculum

The outcome-based curriculum in District 214 has been defined, documented, and structured around two types of outcomes. Specifically, program outcomes have been identified that define what we expect all students to know or be able to do in specific content areas. In addition, ability-based outcomes define what we expect all students to do with what they know. We have identified these outcomes as the general learner outcomes that our students must demonstrate. Consequently, the curriculum is based on the combination of content-based and ability-based outcomes we believe will best prepare students for the challenges they will face after graduation. The following is a brief description of the program outcomes, the general learner outcomes, and the curriculum mapping functions of the IRN.

### Program Outcomes

The district has defined essential outcomes in language arts, fine arts, mathematics, science, social science, physical education/health/driver education, practical arts, foreign language, student activities, and student services. Note the inclusion of our student activities and student services programs as instructional programs: We believe that we must address the totality of the high school experience in the outcome development process, and that these programs needed an appropriate place in our curriculum. Also, Illinois has developed specific goals and standards for learning, assessments, and school improvement mandates in the first six listed programs that local school districts must implement in each school. The outcomes in each of the program areas listed previously have been defined and organized as shown in Figure 2.

Figure 3 shows an example of the relationship among program goals, outcomes, and performance-based indicators drawn from our documentation of the district's science program outcomes. Program goals, outcomes, and indicators of achievement are further placed into courses where all students have opportunities to achieve each of the program outcomes.

### General Learner Outcomes

In 1995, District 214 began requiring graduates to achieve 11 interdisciplinary General Learner Outcomes. These outcomes are the result of an ongoing commitment to define what competent graduates can demonstrate as a result of their learning. They include

- Analytic capabilities
- Civic responsibility
- Ability to communicate
- Skill in creative expression and in responding to the creative work of others
- Responsible participation in the global environment
- Skill in life and career planning
- Problem solving skills
- Facility in social interaction
- Skill in using technology as a tool for learning and expression
- Skill in making value judgments and decisions
- Skill in developing and maintaining wellness.

Figure 2

HIERARCH OF GOALS, OUTCOMES, AND INDICATORS OF ACHIEVEMENT

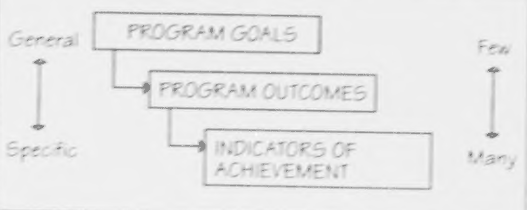


Figure 3

EXAMPLE OF RELATIONSHIP AMONG PROGRAM GOALS, OUTCOMES, AND PERFORMANCE-BASED INDICATORS

Students will understand the concepts of the biological, physical, and environmental sciences and apply them to life in contemporary society.

Students will demonstrate their understanding of transformation, transfer, and conservation of energy by

- Describing the flow of energy through biological systems, e.g., photosynthesis, respiration, food webs, and energy pyramids.
- Describing transfer of energy in physical systems, e.g., rising and falling bodies, periodic motion, simple machines, heat flow, waves, and electrical current.

### The Outcome-Based Curriculum and the Instructional Resource Network

The IRN is designed to store both the content-based and the ability-based components of the outcome-based curriculum. The IRN stores each of the 11 General Learner Outcomes and allows for the attachment to any program, course or learning experience.

Once the outcomes have been entered into the IRN, retrieving and printing any part of the curriculum is possible. For everyday classroom work, teachers would most often print the General Learner Outcomes or the goals, outcomes, and indicators for a selected course.

The IRN provides a simple word processor for entering the outcomes, which serve as the basis for the curriculum. Although changes can be made as the curriculum undergoes modification, changing the wording of the curriculum is only a small part of the work necessary to maintain an integrated system such as the IRN. In addition, maintaining alignment between curriculum and instructional-strategies materials and learning activities is important.

## Outcome-Based Instructional Design

While the "object" of an outcome-based instructional system is the set of learning outcomes students must achieve, the "verb" is the teaching-for-learning process. Without a comprehensive, outcome-based instructional model, the learning outcomes can simply become another version of curriculum documentation that is cataloged and placed on a shelf. An outcome-based instructional design translates the learner outcomes into action plans for both teachers and students.

### Creating School Work Students Can and Will Do.

In addition to aligning teaching strategies and learning activities with learning outcomes, the District 214 vision includes creating school work that students can and will do. The district is committed to designing quality school work that will hold the greatest promise of helping students successfully achieve the intended learning outcomes. Moreover, by stating that school work must be designed so that students can and will do it, two additional criteria emerge. First, most students can complete their school work when we create work that meets their developmental learning needs. By defining prerequisites for each program and course, and clarifying assumptions about students' learning histories, developing instructional strategies sensitive to students' developmental needs and stages-of-learning that are most appropriate for them is possible. Second, school work that students will do can be created by developing engaging, thought-provoking, and relevant learning activities.

### Outcome-Based Instructional Design and the Instructional Resource Network

Anything that can be entered into a typical computer can be entered as instructional material. Most of the material stored in the IRN is produced in Microsoft Word and then imported into the network. In this way, all the features of the computer software can be used to produce high-quality documents.

In the process of designing the IRN, a format for entering instructional material into Microsoft Word allows the attachment of the material to a program, a course, or any level of the ability-based outcomes (general learner outcomes). Once transferred to the IRN, a teacher can search and retrieve any material. Because of the structure of the curriculum, material attached at more specific levels can be retrieved at a higher level. For instance, instructional material attached to a particular course outcome can be retrieved as part of all material belonging to the corresponding course goal, course, or program. Material attached to a General Learner Outcome can be retrieved across program areas. In this way, material appropriate in more than one subject area can be easily shared by others.

## Outcome-Based Assessment System

In developing an outcome-based educational system, consideration of issues related to quality student assessment is essential in evaluating the extent to which students know or are able to do something after instruction has occurred. Current research in cognitive processing, types of intelligence, and learning styles provides a basis for redesigning assessments that use different assessment methodologies to facilitate student learning. Additionally, assessing an ability-based curriculum often requires different types of assessment strategies uncommon in some academic programs. The need for expanded assessment strategies means additional teacher work in terms of assessment development, scoring, and record keeping.

### Guiding Principles and Assumptions

The work in assessment in Township High School District 214 is guided by a philosophy consonant with the principles of outcome-based education. Assessment and testing are seen as essential to a quality education; they serve as "means to an end"—student learning—rather than an end in itself. Assessment should be frequent, purposeful, appropriate, fair, and accompanied by timely feedback. Students must be able to make good use of assessment data in their decision-making processes. Parents also need to have an understanding of assessment data, along with its appropriate use. Teachers need to be experienced in designing a variety of assessment strategies appropriate for the learning outcomes being taught. The assessment strategies should also be appropriate for individual students and enable them to produce the best demonstration of their knowledge and skills. Given different teacher strengths, the sharing of assessment strategies among teachers is important in developing a comprehensive portfolio of assessment strategies.

### Types of Assessments

A variety of assessment instruments is needed to support content-based and ability-based curriculum. The IRN includes

- Test Items
- Close-ended response items that include test items such as matching, true/false, and multiple choice, which have one or more correct answers which are determined in advance.

- Open-ended response items that include test items such as short answer and essay questions. More judgment is required to score these, although qualities of the responses can be determined in advance.
- Performance assessments, which are teacher-designed tasks that ask students to perform in some way. Students are evaluated using a scoring protocol or rubric.
- End-product assessments, which are teacher-designed tasks that ask students to create a product (paper, drawing, art) that can be evaluated using a protocol or rubric.
- Attitude measures, which gather information on students' attitudes (e.g., toward their own learning) and might include surveys, items that measure attitudes, or self-assessments that ask students to reflect on their own performance.

### The Outcome-Based Assessment System and the Instructional Resource Network

In a fashion similar to instructional material, the third section of the IRN allows assessment items entered into Microsoft Word to be imported and stored. Besides an attachment to a part of the curriculum, as before, additional information is also generally stored, such as key words describing what topic(s) an item covers and an item solution. For a multiple-choice item, the solution can be as simple as a letter representing the correct response, while an essay solution will often list key points to be included, and perhaps a system for assigning grades. The solution section is generally not used with performance assessments; instead, a scoring rubric is entered as part of the item to be used by both teacher and student. The IRN can search for assessment items the same way as instructional material based on any of the attached information.

Once common assessment items have been gathered, a teacher selects from them and receives a printout that can be duplicated. If the teacher wishes to create a more traditional test from the close-ended and open-ended questions, a test-making feature of the IRN allows for more involved page layout prior to printing. Features include positioning items on pages to allow for work space, adding page headers, and changing the order of items in the test.

### Modification of Assessments

To allow easy modification of assessment items and instructional materials, an export feature exists in the IRN for transferring material to Microsoft Word. All information attached is exported with the item(s), including curriculum attachments. Once the instructional or assessment materials are transferred, any required changes can be made to the item or the attachments, and the material can be imported back into the IRN. An identification system allows the updated material to replace the original in the IRN.

An important use of the export feature is to change attachments to the curriculum for many items at the same time. One of the difficulties of maintaining a system that integrates instructional material and assessments with curriculum is that as the curriculum changes, many items will need to be attached in a new way. Fortunately, we have found that most materials with a common attachment in the old curriculum continue to have a common alignment with the revised outcomes; therefore, exporting all these items, making the change, and replacing the items in the IRN is possible.

## Analysis of Assessment

The fourth section of the IRN provides for the analysis of data. Initially, this section was designed to analyze multiple-choice tests only. Students would record their answers on mark-sense sheets, and the IRN software would receive the data as the sheets were scanned through a mark-sense reader. Reports could then be printed showing individual student scores, average score for a group of students, and subtest information.

As the emphasis in evaluation has shifted from traditional testing to performance and attitude assessments, this portion of the IRN had to change. A standard mark-sense sheet can be used for inputting rubric results into the software. A teacher marks each student's performance on a mark-sense sheet using a new item for each category of the rubric and the different marks as points. Because the mark-sense sheets in the district use A through E as responses, it has been agreed that A would equal 1 point, B would equal 2, C would equal 3, D would equal 4, and E would equal 5. A five-point rubric covers the majority of those being used in the district, but the software allows for multiple marks to be used on an item, giving a student the sum of the values marked. For instance, marking a B (2) and an E (5) would result in 7 points being assigned. Using this method, up to 15 points can be assigned to each category of the rubric.

At present, the statistics produced for rubrics are identical to those for multiple-choice tests. Although not all the statistics are meaningful for rubric scoring, the majority provide information which, when used properly, can be valuable. Subtests can easily be established consisting of any of the items on a test (or categories of a rubric).

Attitude assessments are often multiple choice in nature. They can be recorded on mark-sense sheets, and the IRN software can read the data. But the IRN software, as originally designed, could give only a small amount of useful information. Rather than making significant changes to the IRN, the decision was made to use the software only for scanning the data.

A data export feature was added to the analysis section of the IRN to place the data into a disk file. The data can then be transferred to a statistical package for further analysis. With the export feature, using an external statistical package to further test data analysis is possible.

## Next Steps

As we reflect on the current stage of the development and implementation of the support functions provided through the IRN, we see a number of additional steps needed to strengthen and enhance the features of the Network. First, we must continue expanding the degree to which the system can accommodate the ongoing renewal of the curriculum, while at the same time maintaining the integrity of the system with respect to the alignment of the curriculum, instructional strategies, and assessments with the learner outcomes. In addition, we continue to encounter a wide range of design questions—for example, we have yet to address the provision of video examples to accompany the description of instructional strategies in the instructional design component of the IRN.

An additional set of issues is the human resources dimension of the support system. While one of the goals of the IRN is to reduce the workload of the faculty in implementing the district's instructional model, a considerable amount of time is still involved in updating and expanding these instructional resources.

Overall, we have found that instructional support services developed and coordinated through the IRN have been instrumental in supporting the development and implementation of our outcome/performance-based system. Nevertheless, as the complexity of the system expands, we must continue to refine the current functions of the IRN and to explore new design features that can support the district's instructional model, if the IRN is truly going to serve a dynamic curriculum and a genuine performance-based assessment system.

# IV-4

# Good Measure: Assessment as a Tool for Educational Reform

DENNIE PALMER WOLF
HARVARD UNIVERSITY

PAUL G. LEMAHIEU
UNIVERSITY OF DELAWARE

JOANNE ERESH
THE CONNELLY CENTER

First published in *Educational Leadership* 49, 8 (May 1992): 8–13. Reprinted with permission of the authors.

© Richard McComb. Kentucky Department of Education

# Good Measure: Assessment as a Tool for Educational Reform

DENNIE PALMER WOLF, PAUL G. LEMAHIEU, AND JOANNE ERESH

*The national call is for improved accountability through high standards and better assessments. What might a school that uses performance tasks and portfolios to achieve these goals look like? Here's one vision.*

Headlines proclaim "National standards and testing." America 2000 urges both. The National Council on Education Standards and Testing has said both are feasible and desirable. Congress has moved forward on legislation, and various professional organizations are collaborating on standards for learning, teaching, and assessment. In short, nearly everywhere the wager is that American education can be galvanized by setting high standards and using new, more probing assessments to hold districts, teachers, and students accountable.

The many projects in this area — whether federal, state, or independent initiatives — share certain commonalities. First, the call is for all national disciplinary organizations to articulate clear, high standards for what students should know (content standards), and how well they should know it (performance standards). Second, there is a push for individual states to establish delivery standards — a description of what all schools must provide so that all students have a fair chance of achieving these standards. At the same time, it is suggested that the achievement of all students, not just a sample, be assessed at 4th, 8th, and either 10th or 12th grade. It is argued that if such assessment is to affect student or system performance, it must be increasingly performance-based. That is, rather than answering a hundred multiple-choice items about American history, a student might be asked to write an essay about the origins and consequences of the Harlem Renaissance based on readings and oral histories of the period. Finally, these standards and assessments are to come together in a voluntary national system.

Albeit the urgency to join and line up is likely to distort volunteerism, ideally individual educational units (districts, states, coalitions of schools) will remain in charge of their own curriculums and assessments, which would be reviewed, coordinated, and certified by one or more overseeing organizations. However, such independence — given the costs — could be quite vulnerable. Still, the wager is that gains will outweigh losses. The hope is that such standards and the pressure to do well in assessments will produce a new level of accountability in education, driving schools to deepen and equalize what they offer.

But at least two serious hitches threaten this wager. First, our history is thin when it comes to standard setting and assessment. We know how to design basic skills testing; how to use test data to rank, rather than improve, schools and to sort, rather than educate, children. We have rarely developed productive, rather than reductive or punitive, assessment and accountability systems — despite the fact that our students are among the most tested in the world. So this bet is based on the *hope* that we can overcome that history and turn standards and testing into productive tools to guide reform.

Second, as they say, "God is in the details." Ultimately, the realization of these national policies on standards and assessment will be played out not in Washington but in the way a teacher answers a child's question; the way a principal finds joint planning time for teachers; and the way a community of parents and teachers joins forces to see that these standards become a tool for, rather than an obstacle to, equity. It depends, as well, on our ability to design and live by frank reporting and consequential learning from assessments.

It is insufficient to declare a broad national policy calling for high standards of performance for all students without an accompanying vision (or better yet, several visions) of what it would be like to *realize* high standards, wise assessment, and thoughtful forms of accountability. What follows is an interim vision pieced together from working closely with teachers and students in classrooms, so it is at least plausible.[1] We have stolen parts from all levels of education, but put it together as a model for middle schools.

**We have rarely developed productive assessment and accountability systems — despite the fact that our students are among the most tested in the world.**

### A Culture of High Expectations

Imagine you are Mr. Rudder, a middle school mathematics teacher who knows how critical it is for students to enter demanding high school mathematics classes if they want to go on to college or decent jobs. You also know how rare this is for many of the students you teach, and you know it is not enough simply to enroll everyone in pre-algebra and demand low failure rates. As a result, you are trying to invent and sustain an enabling mathematics culture that insists on demanding content and high standards for all students. You also have to find ways to make those standards both public and attainable. How is this possible?

**You want to know what kinds of authors, investigators, and speakers your school is turning out.**

To begin with, all your students — those who would have been in pre-algebra and those who would have been in general math — work on open-ended mathematics projects that start with focused problems and move into larger and more complex problems that demand a variety of mathematical skills. Second, the going rules are that each nine-week grading period, every student has to do every project, five out of five must be passing, but two of the five must be better than satisfactory. To make this possible, you and your students discuss what the standards for good work are. In fact, you write them on butcher paper and tape them to the wall. Listed up there are not terms like "neat" and "correct," but "solves the problem" and "shows clearly how the thinking was done." As students realize new dimensions of good work, they add, cross out, and edit.

Unlike usual assignments, there are wider time limits; students keep working and submitting their work for suggestions and critique. Unlike the traditional system in which the unit on percents slams shut with a final test, you encourage students to return to earlier projects with understandings acquired later. In addition, part of what students are appraised on is their ability to find good help. Projects include a list of people a student has conferred with. Moreover, one independent project that a student can choose each nine weeks is to teach another student something mathematically important. Finally, while the specifics of these projects change from year to year, many of the topics, or big ideas, remain common (for example, probability, ratios, graphing). The classroom contains binders of student work from earlier years. As students work, they consult these binders as important resources. There they can see other students' step-by-step thinking and realize that there are multiple solutions to a project. To help students achieve the concentration and stretches of work necessary, community volunteers staff an after-school "homework hall" where cross-age tutoring is the mainstay.

You're not bothered by the fact that some students will take more time or need lots of help. "The standards are the same for all of my students," you say. "All that differs is how we achieve them."

When asked, "Isn't this a lot of work?" you answer, "No question, but the point is my students *have* to have the experience of doing quality work if they are to understand what that looks and feels like. Sort of like learning to ride a bicycle; if you never get on, what are the chances you ever really learn to ride?"

### Internal Accountability: Portfolios and Exhibitions

There is no question that schools owe their constituencies (legislatures, families, and the businesses who will hire their students) honest accounts of what they have and have not achieved. Even so, assessment has been driven too exclusively by concerns for measuring and reporting achievement data for outside audiences. Often forgotten is the equally important work of *internal accountability* — that is, encouraging students, teachers, and families to think hard about what is worth knowing and then making sure students know it.

Think of yourself as Mrs. Fogel, principal of a middle school, designing a set of practices that would make this kind of internal accountability possible. You want the full school community — students, teachers, parents, and possibly even interested outsiders — to be able to look at who is learning what. That desire has several implications. First, you want to know what kinds of authors, investigators, and speakers your school is turning out — not what kinds of test-takers or blank-fillers. Second, you want to know who can perform at a high standard and who the school is and isn't helping to move toward that standard. Finally, you want to be able to see enough of the real (as opposed to the official) curriculum to understand why your results look as they do. Given all of this, you have decided to move to assessing student work — as well as the work of teachers and the school as an educational organization — by performance tasks and portfolios of student work. This decision has brought major changes for students, teachers, and administrators.

To begin, students in your school have to become responsible partners in documenting their learning. In every class, a student puts together a folder that contains class work, journals, and projects. At the end of each nine weeks, teachers work with students to select the pieces that provide an honest picture of the work they can accomplish in that subject or in interdisciplinary projects. Teachers suggest what might be revised and how, or what other work to include.

This *initial portfolio* replaces what used to be the quarterly final or the unit test, but ideally it does more. At best, organizing their portfolios allows students to reflect on what is worth documenting — either because the ideas are important, or because the work has come a long way. On the basis of this portfolio, teachers assign grades or write narrative reports. The reports go home with a sample of a student's typical and best work from each major subject or project. Parents are asked to look at this work with their children and write or call in a response. Once "all the data are in," students reflect on what they have learned from these responses, keeping samples of their best and their typical work.

At the end of every semester, there is portfolio week — much as there was once finals week. Students select samples from each class to create a larger *semester portfolio*. This portfolio becomes the basis of a conference with an adult advisor — a homeroom teacher, subject-matter teacher, or guidance counselor. The conference focuses not just on subject-matter achievement, but also on a student's plans, what major learning tasks are still to be accomplished, what courses to take, and whether there are circumstances that make it hard to accomplish what needs to be learned. During this conference, advisors and students write a letter home that accompanies the portfolio and a report card (or a narrative). Students take their work home to share with their families or with another reader the school has helped find.

Bonnie Thoms

## Rites of Passage and Exhibitions

Looking beyond subject areas, you and the teachers on your staff have met and chosen a small set of essential learning tasks that you believe ought to be the outcomes of middle school. They include such things as being able to conduct an investigation; being willing and able to enter the life of a culture different from the student's own; being able to collaborate on a major project; mastering an aspect of technology. A list of these important tasks appears on the walls of your office, and students have copies inside their folders and portfolios. There has also been much talk about the level at which 8th graders ought to be able to carry out these tasks.

At the close of each year, students create a *year-end portfolio*, drawing on all their work to provide the best, most current evidence of their progress on the essential learning tasks. Students compose a portrait of themselves as able learners, selecting and presenting evidence that they have met the learning standards for individual classes and for the broader learning tasks. They present their portfolios at year-end exhibitions, where committees of readers (an advisor, a teacher, another student, and a family member or guardian) can question and discuss their work. An incomplete or unsuccessful portfolio will not retain a child, but it does affect plans for summer work and next year's school program. Students must successfully meet the portfolio demands before going on to high school. The widespread understanding is that it may take some students three years, others three and one-half years or even four.

As a middle school faculty, you and your teachers have collaborated with the guidance staff and faculty at the high schools you feed. A lean and efficiently notated version of a student's portfolio is the basis for matriculation into high school courses. Your aim is to see increasing numbers of your students go into algebra, honors English, foreign language, and demanding science courses.

But portfolios are not designed just to produce report cards or guide passage into the next year. They have other real and interesting consequences. For instance, at an open house at the end of the year, students draw on their portfolios to make presentations to incoming students about what the newcomers will be learning. Also, portfolios are the basis for participating in special activities. If a writer-in-residence is coming to teach a class, a student can submit writing samples along with an application. Samples of math work can help a student get a community service job tutoring younger children. In these varied ways, students see that their effort counts.

### Helping Teachers Become Judges

The effort that students put into their portfolios is only part of the picture. Teachers, too, find themselves faced with new kinds of work. Matching and true-false tests are worlds easier to correct than are graphs, poems, and models. Moreover, with projects and portfolios, teachers have to grapple with some of the hardest judgment calls. How much should effort count? How does one recognize the scientific reasoning in the writing of a child who is just learning English?

As a principal, you realize the demands portfolios will make on teachers. You create common planning time for teachers who are trying out social studies or mathematics portfolios. A full year's inservice workshops provide occasions for thinking about, responding to, and rating portfolio work. As often as not, there is no outside "expert." Instead, teachers present student portfolios for discussion with their colleagues, who act as a critical review board. A major and sustained focus of this work is the development of *rubrics*, carefully articulated scoring systems, accompanied by samples that typify less-than-satisfactory, satisfactory, strong, and excellent work in subject areas and in the large learning tasks.

In addition, your supervision and work with teachers includes portfolios. Before observing a class, you often request a sample of portfolios to gain a picture of the real curriculum, responses to student work, and levels of achievement. Realizing the success of this at the building level, you have urged the district to work on similar kinds of supervision and mid-career training. Finally, you have become outright "greedy" for your teachers, searching for training opportunities that will let them hone their skills. You have sent some to writing projects or summer institutes, and you have scraped together money to pay three of the most avid portfolio teachers to spend a week of their summer writing up their ideas.

### Community Accountability

Portfolio work comes with no guarantees, no nationally normed scores or grade-level equivalents. How is anyone to know whether it is a hoax or a route to achievement? Some of the schools in the district have made major changes in their curriculum toward sustained projects; others have petitioned to have district tests waived. How are families to know whether children are being cheated or standards lowered because of "some new educational fads"?

First, portfolios do not operate solo. Teachers and the office of testing and evalution are working toward using portfolios to chart *growth* in individuals or groups of students. The data are eventually to be reported in concert with a core set of performance assessments common across schools.

Working together, the portfolio schools have invented additional solutions. Once a quarter, for example, teachers *cross-read*. The science teacher at one school reads a small selection of portfolios from a second school. Language arts and social studies teachers also exchange student work. Time is precious, so they exchange just four portfolios: one from a superstar, one from an average student, one from a student who is struggling, and one that is puzzling. They act as tough critics for one another, asking whether the curriculum is challenging enough, whether teachers are providing enough commentary to students, and whether substantial learning appears to be occurring across the broad spectrum of students. Their rubrics guide this work.

Once a year, a panel of teachers from receiving high schools conducts an *external reading*. They look at a sample of portfolios in their subject from a range of students and comment

**Students compose a portrait of themselves as able learners, selecting and presenting evidence that they have met the learning standards for individual classes and for the broader learning tasks.**

on the curriculum and the quality of student work. Much of the back-and-forth swirls around what changes middle school teachers could make. But middle school teachers also press high school teachers to become involved in placement, to look at portfolios when making up class rosters, and to blur or do away with tracking.

Several times a year, a similar event is held involving a communitywide *audit committee*, composed of business people, superintendents, principals, strong teachers from neighboring districts, and several outside experts. Some students present sample work from their portfolios. Committee members subsequently examine a sample of portfolios. At one meeting, the portfolios might come from a particular subject area; at another, the committee might examine students' year-end portfolios that incorporate materials across subject areas. They look at the quality of work, skill of teacher responses, appropriateness of grades, and whether all students have access to learning. The committee has three major purposes: (1) to publish a report of their findings and the implications of those data; (2) to design a long-term plan for how the district might move responsibly away from standardized testing; and (3) to think through how portfolio and performance task data can be responsibly used to inform planning, budgeting, and school change.

The report is reviewed with collaborating teachers and then presented at an open meeting to which the wider community of teachers and families is invited. The meeting is clearly about choices and values. Sometimes feeling runs high, and the questions from the audience are tough. However, the meeting brings into the open the questions of what is worth knowing and who is learning it. It is accountability at its most immediate — and perhaps its most accurate — level.

## Some Last Thoughts

This is not easy or short work. It is wearing, in part because it is being invented, one step after another. It is also schizophrenic: how does a principal or a parent make sense of portfolio reports, performance scores, and the safety-net standardized testing the district still does? When portfolio reports and test scores conflict, which one should be believed? It is also hard because it flushes fundamental values to the surface. One English teacher believes that we mislead children, even handicap them, if we ever let them write in nonstandard English. Another believes that children should be allowed to write in whatever language they can be most fluent and thoughtful in, with corrections coming later. They disagree fiercely and in public.

Parents worry that portfolio work will not prepare their children for the "real world" of high school, college entrance exams, and other tests. There is a question about whether receiving high schools will honor portfolios when they are accustomed to transcripts and test scores. Staff in testing and evaluation know that alternative forms of assessment are in their infancy. And some have real doubts. What do we really know about validity? About gender and racial equity in portfolios?

So the challenges are numerous. However, if national standards and testing are to be educational tools for reform, then all students, teachers, and schools must have ongoing experience with demanding tasks and high standards. Not once a year in April — but continuously. We have to fill the gap between declaring and realizing high standards — not with promises — but with the details of strong educational practice.

---

[1] Much of what is described in this article has emerged from a multiyear project in school reform funded by the Rockefeller Foundation. The project is a joint endeavor of Performance Assessment Collaboratives for Education (PACE) at the Harvard Graduate School of Education and the school districts of East Irondequoit, N.Y.; Rochester, N.Y.; Pittsburgh; San Diego; San Francisco; and Shutesbury, Mass. These schools, in turn, have drawn on assessment practices designed in other schools and programs including Central Park East Secondary School and ARTS PROPEL. A number of papers and profiles of the participating schools can be obtained by writing to PACE.

# IV-5

# When Teachers Look at Student Work

CHRISTINE SOBRAY EVANS
BERNARDO HEIGHTS MIDDLE SCHOOL, SAN DIEGO, CALIFORNIA

When I started teaching 17 years ago, I would look at student work only to give it a grade and check it off. I relied on publishers to test my students and to tell me what to teach. I would start at Chapter 1 and hope to finish the text before the end of June. In those days, my focus was on me—my plans, my lessons, my teaching. It embarrasses me now to confess this. I would like to call all of my early students and apologize for my selfishness.

Since that time, a revolution has taken place in my classroom and in classrooms like mine across the state. In California, we have moved beyond a skill-based curriculum, which has to do with teaching, to a meaning-centered curriculum, which has to do with learning. Because of this shift, student work has become the center of education. Teachers no longer follow the guidelines of publishers but, instead, look at their students to determine what must happen next.

## The Power of Portfolios

A portfolio is one strategy teachers can use to focus on student work. In Poway, we've been using Language Arts portfolios for almost three years now. That might seem like a long time in education, but we are just beginning to reach new understandings.

At first, we coped with the physical problems of portfolios. Who should keep them? Where to keep them? How often should we look at them? What should go inside? Since our new report card stated that "grades" were based on a collection of work, portfolios appeared in some form in almost every elementary classroom.

That first year, portfolios were much like the work folders that teachers had traditionally kept for students. Teachers, rather than students, made most of the selections. The portfolios were organized storage bins, where work went in but never came out. Then at the end of the year, the portfolios were sent home—much like a scrapbook for the year.

In a few places, however, some exciting things were done with portfolios. At Los Penasquitos Elementary School, the 3rd grade team met in a tiny storage room—portfolios in hand—to look more closely at the 3rd grade language arts program. All the portfolios were shuffled and dealt out like a deck of cards. We then spent about 45 minutes, silently reading and making notes. The first surprise was how

---

First published in *Educational Leadership* 50, 5 (February 1993): 71–72. Reprinted with permission from the author.

accurately the portfolios portrayed our students. I remember reading through one folder and saying, "Mary, Sarah appears to be struggling with writing. All of her pieces are short and done just to complete the assignment. Her reading list shows that she enjoys reading grade-level books."

"That's right," Mary replied, surprised that I knew so much about a student I had never met. This interchange was repeated time and time again. It gave us great confidence that portfolios could provide an undistorted image of student effort and achievement.

When we looked at the collection of portfolios as a whole, we made some important discoveries, too. Worksheets and formula writings were much less informative than student-generated topics. Our 3rd grade team decided to spend less time on formula writing assignments that asked students to fill in the gaps—and more time teaching young writers to focus on one topic. So, although we were really looking at *student* learning, we learned a tremendous amount about our students and our 3rd grade program.

### Developing Models and Rubrics

The power of looking directly at student work as a *team* cannot be overstated. Real student work gives teachers a starting point for conversations that get to the essence of what happens in classrooms. Samples of student work are concrete demonstrations of what is known and what is not known. They also provide teachers with signposts that mark how far we've come and point us in the direction we must follow.

At Los Penasquitos, the principal provided released time for the grade level teams to get together. However, in schools where there is staff interest, teachers can still meet. Bernardo Heights Middle School is a perfect example. Last fall teachers on the 7th and 8th grade teams decided to establish schoolwide standards in writing. Using a prompt from the California Assessment Program (CAP) handbook, we asked every 7th and 8th grader to write about a school-related experience. Students completed these essays during a 45-minute class period—all were first draft pieces.

That afternoon we met to select examples of student writing to serve as models for our 6-point rubric. Papers were shuffled about from stack to stack. We placed pieces that showed promise in the *high* pile, papers that looked like typical responses in the *middle* pile, and papers that raised concerns into the *lower* pile.

In the high pile, we hoped to find "outstanding" examples that would rate a 6 and "good" ones that would rate a 5. On the chalkboard, we listed characteristics that put a paper into the top category: selecting a dramatic topic, using precise vocabulary, and creating sentence variety, to name a few.

"I think this might be a 5," said Kathy. She read it to our tiny group and then argued for it: "It has more sophisticated sentence structure than we normally see. The vocabulary is strong, and the details paint a clear picture. I know it's not a 'wow,' but it's clearly a 5."

Our dialogue continued until we had models for each point and had annotated each piece with information for future reference. Each teacher in the group received copies of the models and annotations, and we've shared them with students and parents. We wanted everyone to have a clear understanding of how we look at student writing.

### Spotlight on Students' Work

To our surprise, the models were just one benefit of this entire process. More important, we felt, were the discussion and negotiation that occurred that afternoon. In addition to learning about student writing, we valued the opportunity to talk to one another about the work we do. That sense of being a community of learners continued for the entire year.

Our "community of learners" has since expanded to students and their parents. We use our models and rubrics to train young writers to examine their own work and that of their peers:

- Can you find examples of precise vocabulary?
- Can you play a videotape in your head and see the action?
- Here is a paper written last year. You said it was a 3. Can you work with your group to make it a 4?

Questions like these help us to zero in on the elements that make writing come alive for the reader.

Parents have shown their support by attending evening workshops so that they, too, can learn to look at student work through a new lens. "This is not easy," one mom wrote after a training session on rubrics. "However, I think that I can help my daughter better now."

Teachers, students, parents—we are all focusing on real work to give us insights about learning. Sometimes, though, these insights make us uncomfortable. "I'm going to have to re-do mine," Tom confided to me, "I just didn't know what *good* looked like!"

The portfolios that fill our crates and cabinets, and sometimes cover my dining room table, are no longer scrapbooks of completed assignments. Instead they are showcases for the best work a student can do. They allow young people, like Tom, a chance to reflect, revise, and re-start, if necessary, in order to paint a portrait of themselves as learners.

Portfolios also give me, the teacher, the same opportunity to reflect, revise, and re-start. I no longer look to publishers to tell me what to do next. My students show me the way.

IV-8

# Integrating Assessment and Instruction: Continuous Monitoring

JACKIE PALMER
SOUTHWEST EDUCATIONAL DEVELOPMENT LABORATORY

## Why Integrate Assessment and Instruction?

> From today's cognitive perspective, meaningful learning is reflective, constructive, and self-regulated. . . . People do not merely record factual information but create their own unique understandings of the world—their own knowledge structures. To know something is not just to passively receive information, but to interpret it and incorporate it into one's prior knowledge. In addition, we now recognize the importance of knowing not just how to perform, but also when to perform and how to adapt that performance to new situations. The presence or absence of discrete bits of information, which is typically the focus of many traditional multiple-choice tests, is not of primary importance in the assessment of meaningful learning. Instead, we care more about how and whether students organize, structure, and use that information in context to solve problems (Herman, Aschbacher, and Winters 1992, pp. 14-15)

Based on the cognitive psychology of how learning takes place in individuals, it is imperative for the teacher to monitor the child constantly in order to meet the needs of that child. Spector (1993) describes the new instructional paradigm as one that is generative (guided by student questioning), inclusive (integrating multiple cultural, racial, and gender values) and constructivist (building on student's biological and experiential prior knowledge). It mirrors good science, establishes links, and supports active, experiential learning. It fosters collaboration and takes into account individual learning styles and stages of cognitive development.

Similarly, the role of assessment in the new paradigm must also shift. Rather than rewarding only correct responses, assessment must inquire about the reasons for obtaining incorrect responses (which may not, on reflection, be incorrect at all, if viewed from the student's perspective). The role shift emphasizes different kinds of assessment for different pur-

---

Adapted from Chapter 4 of *Improving Science and Mathematics Education/A Tool Kit for Professional Developers: Alternative Assessment* (Pilot Test Version), September 1993, 4.1-4.8. Developed by the Laboratory Network Program—Alternative Assessment. Available from Northwest Regional Educational Laboratory, Portland, Oregon.

poses, and multiple forms of assessment for determining student learning. Assessment is geared to improving instruction, to stimulating inquiry and personal growth in students, and to fostering cooperative learning. In short, it must assess what society values, and in the educational arena, that is student learning.

The ideal approach for constant monitoring of student learning is the integration of assessment and instruction. In its ultimate form, this integration is so complete that the lines between assessment and instruction fade. The student is completely unaware of being assessed, of instruction being modified on the spot, and of further cycles of assessment/ instruction/ modification of instruction/reassessment, ad infinitum. In the new educational paradigm, then, the integration of assessment and instruction can be viewed as a spiral. Figure 1, based on a constructivist model of learning, depicts this spiral approach.

In this model, assessment is an ongoing activity in the instruction cycle (process) designed to create an optimal learning situation for students. It results in continual evaluation and adjustment by the teacher. For example, when students are asked to explain how they arrived at their answer to a problem calling for a comparison of the area of a field to its perimeter, the teacher is able to assess student understanding of the concepts of area and perimeter, ratio, computation, and non-mathematical skills such as reading and comprehension. If a student is not successful in solving the problem because of misconceptions about the meaning of ratio, the teacher can provide experiences that enable the student to master that concept. The teacher might also determine through this informal assessment that students have already mastered a concept that was planned for instruction in the future, such as symmetry, and eliminate that instruction.

As this cycle of assessment/instruction is conducted by the teacher, adjustments are made that reflect information gained, learning is reassessed, and further adjustments are made as required by new assessment information. In the role of teacher-practitioner, the teacher becomes more knowledgeable about pedagogy and classroom application, which results in growth of the individual teacher to deliver appropriate instruction to all children. The assessment becomes, therefore, not an end in itself, but a tool for discerning the need for change.

Effective instruction must be integrated with assessment. To fail to do so means to continue to lose from the educational pipeline those students who have traditionally not performed well in the old educational paradigm, e.g., those "at risk." *Not* integrating assessment and instruction substantially hinders progress towards achieving the third National Educational Goal put forth by President Bush and the National Governors Association: ". . . every school in America will ensure that *all* students learn to use their minds well, so they may be prepared for responsible citizenship, further learning, and productive employment in our modern economy" (italics added) (*America 2000* 1991).

## What Is Integration of Assessment and Instruction When Viewed as Continuous Monitoring?

In general, newer science and mathematics curriculum materials advocate assessments that:

- Are embedded within instructional materials
- Use a variety of methods to assess the student's progress
- Emphasize teacher observation and teacher judgment
- Provide methods for getting at the reasons behind children's answers (Perrone 1991, p. 18).

Having established the efficacy of integrating assessment and instruction, there is still much confusion about what integrated assessment and instruction looks like. There are basically two key characteristics of integrated assessment and instruction: first, integrated assessment must be a process of inquiry by the teacher, and second, the assessment must impact the instructional process and thus contribute to learning.

### Continuous Assessment of Student Learning/Developmental Growth

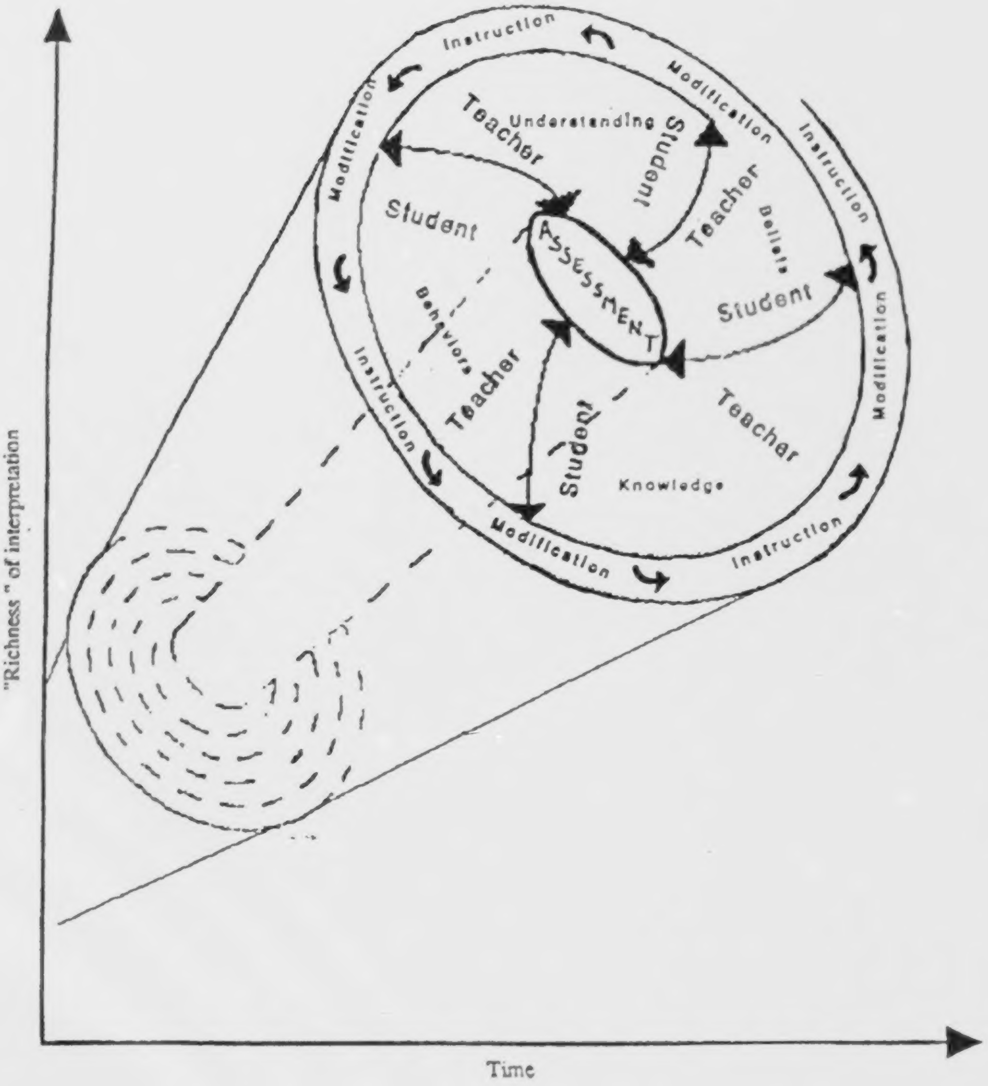
In order to get a more concrete view of successfully integrated assessment and instruction, it is helpful to review the following list of assessment methods modified from Perrone (1991):

- Observation
- Verbal response (interviews, oral examinations, informal oral questioning)
- Written records, narrative form (essays, stories, descriptions of processes)
- Written records, drawings/illustrations (diagrams, conceptual maps and webs, graphs, sketches, detailed drawings)
- Demonstration of skills (dramatic responses, successful experiments/use of manipulatives)
- Products (models, displays, projects, dramatic presentations).

Note that many of these methods can be used in tandem. For instance, the "student-as-teacher" approach to integrating assessment and instruction may utilize all of the above approaches. In one integrated assessment/instruction task developed by a high school teacher, for example, senior physics students use an inflatable portable planetarium to

Figure 1

A SLICE OF TIME IN A CONSTRUCTIVIST CONTINUUM OF CURRICULUM DEVELOPMENT


Instruction
Modification
Modification
Instruction
Teacher
Understanding
Student
Student
Teacher
Beliefs
ASSESSMENT
Student
Behaviors
Teacher
Student
Teacher
Knowledge
Instruction
Modification
Modification
Instruction
"Richness" of interpretation
Time

BEST COPY AVAILABLE

learn and teach basic astronomy facts. In pairs, they research and prepare an astronomy teaching presentation for one or more science classes in the school. Few teacher-developed tests are used to determine astronomy content knowledge in order to reteach concepts so that students feel comfortable teaching them. Using various teacher-developed rubrics, students are continually assessed on their research and collaborative skills. During planetarium productions, they are further assessed on their responses to student questions and adaptations of their presentation to the audience—skills particularly enhanced by content knowledge and understanding.

In another student-as-teacher approach that also combines self- and peer assessment, a team of high school science teachers developed an integrated assessment and instruction activity in which pairs of students prepare chemistry demonstrations focusing on selected biological and chemical processes. They then perform the demonstrations for younger students in other science classes, responding to their questions, while class members videotape the productions. Class members subsequently critique their own and each other's performances using the videotapes and criteria developed jointly by the students and teacher.

Figure 2, an excerpt from a paper on Cognitively Guided Instruction (Fennema, Carpenter and Petersen, undated), further demonstrates how assessment and instruction can be effectively integrated in mathematics.

Note in this example that the assessment is a continual process aimed at identifying the student's cognitive level so that instruction can be modified accordingly. Without the continuous assessment, the instructor would be ignorant of whether or not students were cognitively ready to move on to the next stage of understanding. The Cognitively Guided Instruction Model thus has both of the key characteristics of integrated assessment and instruction—it is a circular process of inquiry used to modify instruction.

The Mathematical Sciences Education Board, in the text *Measuring Up: Prototypes for Mathematics Assessment* (1993) provides prototypes of integrated instructional/assessment activities that support the NCTM Standards. In the prototype "Bridges," elementary students are provided with manipulatives and student worksheets with the following instructions:

> You are an engineer who wants to build different kinds of bridges. The bridges will be made of colored rods. The first bridge you are to build is a 1-span bridge made with one yellow rod and two red rods . . . The yellow rod is called a span and the red rods are called supports. Since the yellow rod is 5 cm long, the length of the bridge is 5 cm . . . As you build bridges in the following activities, think of a way to keep track of the number of rods of different colors you use. Your goal is to find out how many rods of each color you would need to build a bridge of any size. (*Measuring Up*, 1993, p. 44)

FIGURE 2

### EXCERPT FROM "LEARNING MATHEMATICS WITH UNDERSTANDING"

A major problem in the sequence of development of mature problem solutions is the use of symbols. Consider how Ms. Pink helped her children to develop their own understanding of the complex process of writing number sentences.

> *The children in Ms. Pink's class could solve some kinds of problems quite well. Some children used direct modeling, some used counting on, and a few used derived facts. Some children recalled facts with small numbers. No one used written symbols. Ms. Pink asked a small group to join her near the blackboard and asked them to solve a specific type of problem. Ms. Pink then showed the children how that problem looked when represented by symbols and asked the children to explain the symbols. One of the children gave the following explanation for transferring what she did with her counters to the number sentence. Numbers show the same things the counters show.' Ms. Pink continued to have the children solve problems and to write the number sentences on the board. She asked the children to relate how they had solved the problem to the number sentences . . .*

Ms. Pink had carefully observed her children and decided that this group was ready to move on to using symbols in conjunction with their more primitive problem solving procedures. For the next several months, she monitored carefully their understanding of the relationship between strategies such as direct modeling and the writing of symbols. She never expected children to perform operations with symbols that they could not perform readily by direct modeling or counting strategies. However, she recognized that the most mature strategy that children could use involved symbols, so she continually pointed out the relationship of symbols to their less mature strategies. She believed that the children would eventually construct the relationship between symbols and other strategies and thus come to use the symbols with understanding.

Fennema, Carpenter, and Peterson, undated
(permission to use granted by Fennema)

Students then proceed to work through the activity. They are encouraged to monitor their own understanding and to "decide for themselves what they need to do to understand the situation" (p. 49). The task, while focusing on "the generation and identification of patterns in a numerical setting . . . is designed so that a variety of approaches to the computations are feasible . . . More important than the computation itself, however, are the students' decisions about what computations are appropriate, and the connection between the computation and the dimensions of the bridge" (p. 49). It is these ongoing decisions that must be monitored with ongoing assessment.

### The Advent of New Technologies

The advent of new technologies has the potential to further expand our ability to integrate assessment and instruction. An example of a prototype instrument currently under development is provided in the LNP Alternative Assessment database.[*] The "Multimedia-Based Assessment of Science Process Skills" uses a computer linked to a videodisc player to simultaneously instruct and assess student process skills.[**] Eleven performance tasks are being piloted at both the elementary and middle school levels. In the task "Fox Rot" (see Figures 3, 4, 5, and 6) students observe photos of a fox decomposing, place them in chronological order, and explain reasons for their ordering. Students may then view a short movie sequence as a form of self-evaluation, after which they can modify their responses. Responses may either be elicited verbally by the teacher, or they can be entered into a CLARIS Works program for printing or electronic network access by the teacher. A second task, "Are You Listening?" (see Figures 7 and 8) is an interactive program designed to convey and assess the concepts of pitch and intensity. Tones are simultaneously heard and visually depicted via frequency curves. Students are asked to match tones with curves. They may record sounds in the room, predict what the frequency curve might look like, and test their predictions. In each of these tasks, assessment is by the student and/or teacher, with immediate feedback and modification of instruction readily administered.

From each of the above examples, key words have been highlighted to indicate approaches that have been successfully employed to integrate assessment and instruction as a monitoring tool: student-as-teacher, self assessment, peer assessment, cognitively guided instruction, manipulation of objects to assist in making mental connections, and application of new technologies to integrating assessment and instruction. This is not, by any means, an exclusive list of approaches. Any of the tasks listed at the beginning of this chapter can serve, either singly or in combination, as vehicles for integrating assessment and instruction. They need only be applied with the purpose of inquiring about the status and/or process of student learning, with the intention of modifying instruction based on that assessment, in a continuous fashion.

## CONCLUSION

If we truly want to ensure that *all* students attain the more complex skills and thinking processes that many feel are important to be successful and productive citizens for the future, we must ensure that we know both what success in these complex areas looks like and what satisfactory progress toward the ultimate goal looks like. Attention to the careful design of assessment forces us to confront these definition issues. Viewed in this manner, good assessment becomes not just an exercise in monitoring set apart from instruction, but the essential ingredient that forces us to be crystal clear about what we want to accomplish with our students, and simultaneously, a tool for helping students attain those goals.

> Some researchers and educators look forward to the day when there will be a "seamless web" of instruction and assessment in which assessment is no longer a distinct activity but is 'built into' students' regular learning experiences, virtually indistinguishable from instruction. In this vision, students and teachers receive ongoing feedback as needed during the instructional process. Assessment for classroom purposes (as opposed to external monitoring purposes) is a routine, non-threatening process. This vision calls upon students, teachers, and administrators to look at assessment in a different way. Teachers must understand that the primary information will not be

---

[*] The LNP-AA database is a compendium of descriptive information on 137 alternative assessments in science and mathematics. It is available from Document Reproduction Service, Northwest Regional Educational Laboratory, 101 S.W. Main St., Suite 500, Portland, OR 97204, (503) 275-9519. (NOTE: This database is available for both IBM and Macintosh computers and uses FileMaker Pro software, which much be purchased from Claris.)

[**] A *Multimedia-Based Assessment Sampler* is available from the AOP Hub/Midwestern State University Assessment Consortium. The Sampler includes portions of the most recent field test versions of "The Fox Rot" and "Are You Listening?" along with other tasks under development. The *Sampler* is formatted for Macintosh computers only (4MB RAM, a hard drive, and a color monitor are recommended. Quick Time and Hyper Card 2.1 software are required). To receive the *Sampler,* send a 3.5" high density (1MB or greater) disc and $2.00 for postage and handling to the author (see Reference list at end of this chapter).

> quantitative, but will provide a rich portrait of student strengths and weaknesses. In addition, assessment to inform instruction does not have to compare students to one another, which means that student approaches and responses to a problem may look very different and still be correct . . . [Such assessment] should measure the processes students use as well as the answers they reach. It should measure all of the goals of the curriculum, not just a few. It should address both group activities and individual ones. It should be developed by, or with ample input from, teachers and should include teacher professional observation and judgment. Perhaps most importantly, it should draw upon information from multiple assessment sources, including but certainly not limited to tests . . . Finally, it should have a strong self-evaluation component for both students and teachers. (Kober, 1993 p. 62)

## REFERENCES

*America 2000: An Education Strategy.* (1991). Washington, DC: U.S. Department of Education.

*Curriculum and Evaluation Standards for School Mathematics.* (1989). Reston, Va.: National Council of Teachers of Mathematics.

Erickson, M. (Winter 1992). "Developing Student Confidence to Evaluate Writing." *Quarterly of the National Writing Project & The Center For the Study of Writing and Literacy* 14, 1: 7-9.

Fennema, E., T.P. Carpenter, and P. Peterson. (undated). *Learning Mathematics with Understanding.* Madison, Wisc.: Wisconsin Center for Education Research.

Fennema, E., T.P. Carpenter, and P. Peterson. (undated). *Cognitively Guided Instruction Model.* Madison, Wisc.: Wisconsin Center for Education Research, pp. 36-37

Herman, J. L., P.R. Aschbacher, and L. Winters. (1992). *A Practical Guide to Alternative Assessment.* Los Angeles, Calif.: The Regents of the University of California.

Kober, N. (1993). *EDTALK. What We Know About Science Teaching and Learning.* Washington, DC: Council for Educational Development and Research.

Laboratory Network Program. (1993). *Improving Science and Mathematics Education—A Database and Catalog of Alternative Assessments.* Portland, Oreg.: Northwest Regional Educational Laboratory.

Land, M., and T. Peters. (1994). *Multimedia-Based Assessment Sampler.* Wichita Falls, Texas: Midwestern State University, Division of Education.

Lucas, C. (1990). "Introduction: Writing Portfolios—Changes and Challenges." In *Portfolios in the Writing Classroom,* edited by K. Yancey. Urbana, Ill.: National Council of Teachers of English.

*Measuring Up: Prototypes for Mathematics Assessment.* (1993). Mathematical Sciences Education Board. Washington, DC: National Academy Press.

Penick, J. (November 1989). "Issues-Based Education." *Education and Urban Society* 22, 1: 3-7.

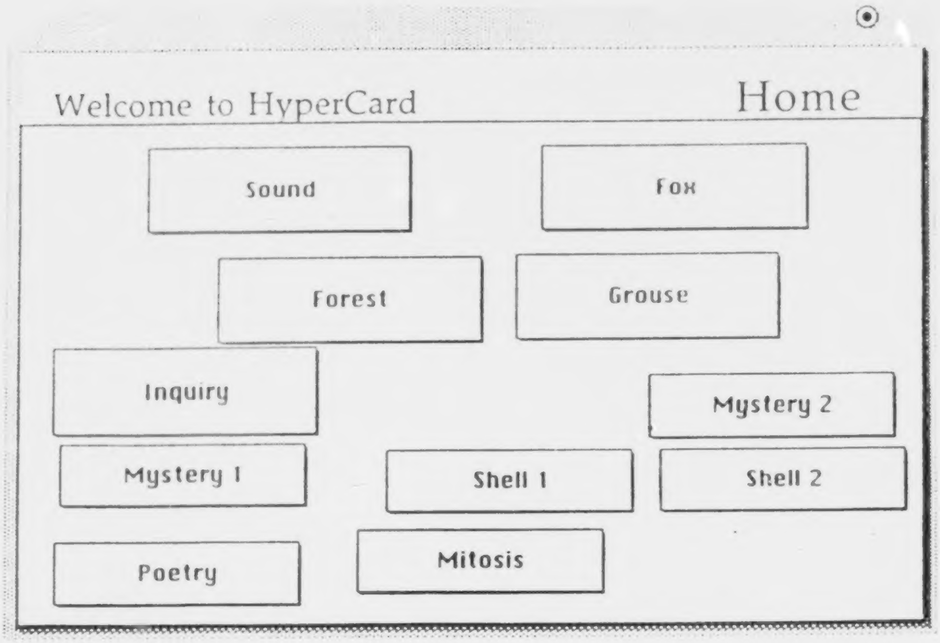
Perrone, V., ed. (1991). *Expanding Student Assessment.* Alexandria, Va.: Association for Supervision and Curriculum Development.

Spector, B.S. (January 1993). "Order Out of Chaos: Restructuring Schooling to Reflect Society's Paradigm Shift." *School Science and Mathematics* 93, 1: 9-19.

Tierney, R.J., M.A. Carter, and L.E. Desai. (1991). *Portfolio Assessment in the Reading-Writing Classroom.* Norwood, Mass.: Christopher Gorden Publishers.

FIGURE 3

HOME CARD AND ASSESSMENT TASK INFORMATION


Welcome to HyperCard
Home
Sound
Fox
Forest
Grouse
Inquiry
Mystery 2
Mystery 1
Shell 1
Shell 2
Poetry
Mitosis

"Non Enriched" Performance Task: an activity designed to assess a narrow range of student achievement which can be completed in a short time frame (one class period or less).

Key Features:

prompt: a charge to student to undertake a task which leads to a performance or product. (Green, 1990?)

rubric: a scoring system enabling the assessor to place value on important aspects of the performance or product. (narrows the range of achievements measured)

Purpose:

formative: setting value on student achievement during an instructional episode. The primary purpose is to direct future instruction.

DOMAIN II - Exploring & Discovering. Processes of Science.
DOMAIN III - Imagining & Creating. Higher Order Thinking/Problem Solving

*Multimedia-Based Assessment Sampler*
*Land and Peters, 1994*

FIGURE 4

## INITIAL PROMPT AND RUBRIC—THE FOX ROT

### Performance Task – The Fox Rot

**Part 1: Field Research**

**PROMPT:** Observe the 4 photos of a fox decomposing. Place the photos in order from start of decomposition (top) to end of decomposition (bottom). List the observations that you make to support the order in which you place each photo. Be sure to compare data from each photo.

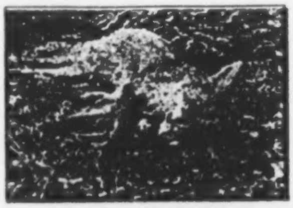
**RUBRIC - Observing/Inferring:**
TX Essential Element 2 (B, C, D) – Acquire data through senses.
TX Essential Element 4 (B, D) – Communicate data and information.
TX Essential Element 6 (B) – Draw logical inferences.

| POINTS | CHARACTERISTICS |
|---|---|
| 4 | Observes/infers properties of the fox (which may include its surrounding environment) for all four photos. |
| 3 | Observes/infers properties of the fox (which may include its surrounding environment) for three photos. |
| 2 | Observes/infers properties of the fox (which may include its surrounding environment) for two photos. |
| 1 | Observes/infers properties of the fox (which may include its surrounding environment) for one photo. |
| 0 | Makes no observations/inferences. |

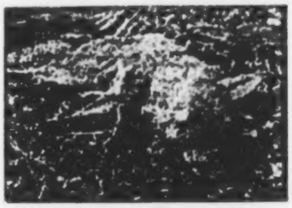
*"The Fox Rot"*
*Multimedia-Based Assessment Sampler*
*Land and Peters, 1994*

FIGURE 5
SAMPLE STUDENT RESPONSE

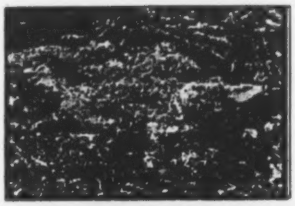
**My Fox Story**



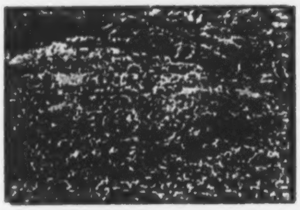
I think that this is the beginning of the fox rotting. It pretty much looks like the fox might be alive but just sleeping. There are no signs of it decomposing yet.



Here you can see how the fox is changing. It is kind of puffed up. I've seen dead animals in the neighborhood and on the side of the road that are all swollen after being dead awhile... especially if its in the summer. All the green plants around the fox make me think it is summer.



Really gross!!! The fox's body is covered in maggots. They must be eating it because alot of the fox is missing. Or, I guess some scavenger might have come along and eaten part of it before the maggots got there.



Its almost all gone now. There's lots more maggots too. If you didn't know it was a fox at the start you almost couldn't tell now.

The End.

*"The Fox Rot"*
*(video footage from Optical Data's*
*"Windows on Science," Texas Edition, 1990*
*Multimedia-Based Assessment Sampler)*
*Land and Peters, 1994*

BEST COPY AVAILABLE

## FIGURE 6
## VERIFICATION PROMPT AND RUBRICS

**Part 2: Verification**

**PROMPT:** View the time lapse sequence of a fox decomposing. You may view it as often as you like. Compare the sequence of photos you made in Part 1 with the time lapse sequence. Does your photo order match the time lapse sequence? Explain your answer using observations to support you conclusion.

**RUBRIC - Conclusions:**
TX Essential Element 4 (F) – interpret the arrangement of data on...visuals.
TX Essential Element 6 (D) – draw conclusions from observed data.

| POINTS | CHARACTERISTICS |
|---|---|
| 4 | Draws conclusion. Bases conclusion on observations comparing all still photos of the fox with the time lapse segment. |
| 3 | Draws conclusion. Bases conclusion on observations comparing three still photos of the fox with the time lapse segment. |
| 2 | Draws conclusion. Bases conclusion on observations comparing two still photos of the fox with the time lapse segment. |
| 1 | Draws conclusion. Bases conclusion on observations comparing one still photo of the fox with the time lapse segment. |
| 0 | Draws no conclusion. |

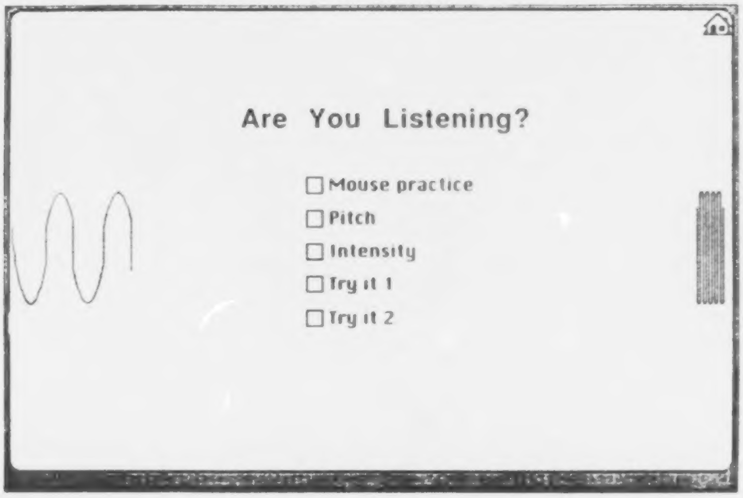
**RUBRIC - Sequencing:**
TX Essential Element 3 (A) – Classify, order, and sequence data.

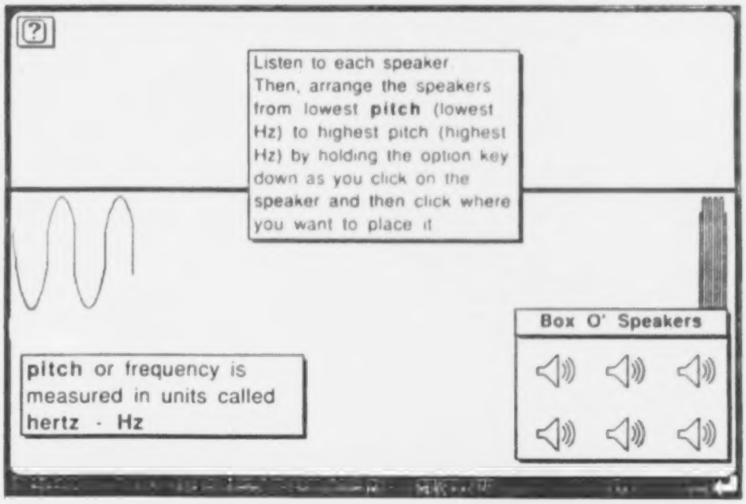
| POINTS | CHARACTERISTICS |
|---|---|
| 3 | Sequences all 4 photos to match actual sequence. |
| 2 | Sequences 2 of 4 photos to match actual sequence. |
| 1 | Sequences 1 of 4 photos to match actual sequence. |
| 0 | Sequences none of 4 photos to match actual sequence. |

*"Are You Listening?"*
*Multimedia-Based Assessment Sampler*
*Land and Peters, 1994*

FIGURE 7

INITIAL SCREEN AND INSTRUCTIONS


Are You Listening?
☐ Mouse practice
☐ Pitch
☐ Intensity
☐ Try it 1
☐ Try it 2


Listen to each speaker. Then, arrange the speakers from lowest **pitch** (lowest Hz) to highest pitch (highest Hz) by holding the option key down as you click on the speaker and then click where you want to place it.

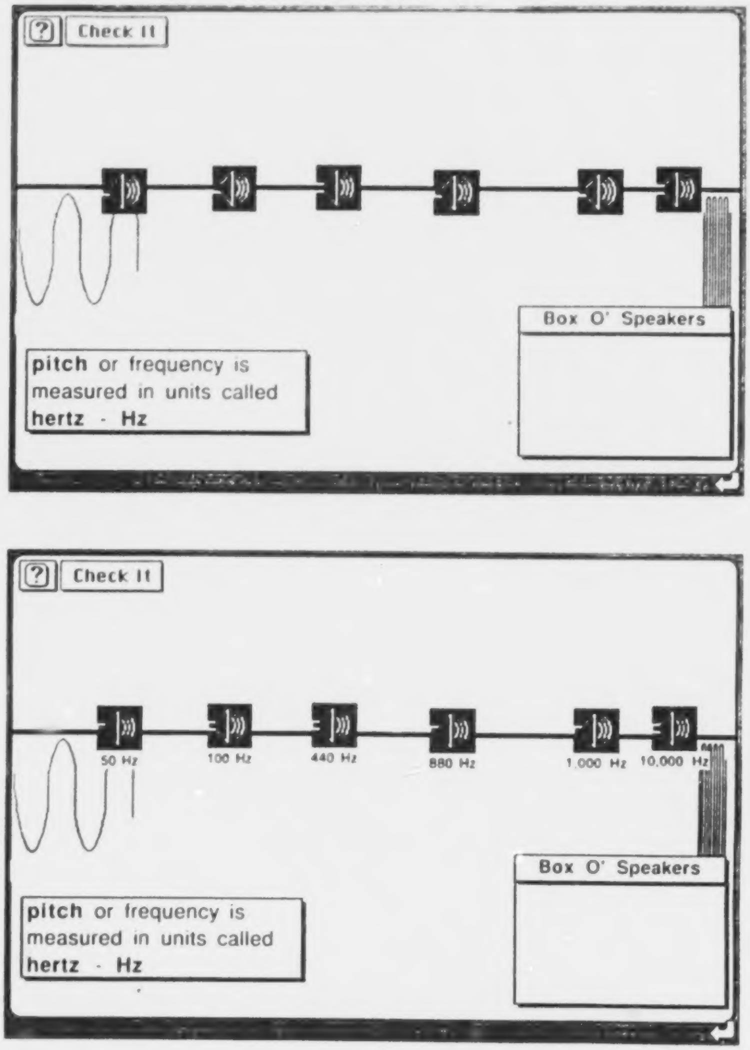
**pitch** or frequency is measured in units called **hertz - Hz**

?
Box O' Speakers

*"Are You Listening?"*
*Multimedia-Based Assessment Sampler*
*Land and Peters, 1994*

FIGURE 8

STUDENT RESPONSE AND VERIFICATION


?
Check It
Box O' Speakers
?
Check It
50 Hz
100 Hz
440 Hz
880 Hz
1,000 Hz
10,000 Hz
Box O' Speakers

**pitch** or frequency is measured in units called **hertz** - **Hz**

**pitch** or frequency is measured in units called **hertz** - **Hz**

*"Are You Listening?"*
*Multimedia-Based Assessment Sampler*
*Land and Peters, 1994*

# IV-7

# Linking Assessment with Instruction

Richard J. Shavelson, Gail P. Baxter
University of California, Santa Barbara

*In the United States, educational reforms of the 1980s sought to improve curricula,* increase high school graduation standards, raise teachers' pay, improve administrative leadership, and hold schools accountable for student achievement (see, for example, National Commission on Excellence in Education, 1983). These reforms have, for the most part, failed (Shanker, 1990): students improved basic skills at the expense of understanding and problem solving (Mulls and Jenkins, 1988). This finding reaffirms what was learned in the 1960s mathematics and science reform: students learn, to a greater or lesser degree, what they are taught (Walker and Schaffarzick, 1974). Moreover, teachers, not surprisingly, teach what they are held accountable for—student achievement on fact-driven multiple-choice tests (Shavelson, Carey, and Webb, 1990; Smith, 1991). The inevitable conclusion is that by directly influencing *what* teachers teach, achievement testing has had a profound effect on the *content* and the *manner* of instruction.

Alternatives to traditional multiple-choice achievement testing are currently being proposed (Wiggins, 1989; Gough, 1989; Nickerson, 1989). They rest on a set of assumptions about effective and responsible teaching that differ significantly from assumptions of the past. Perhaps the most fundamental assumption is that learning is an active, mindful process (Glaser, 1984; Resnick, 1987). In this process, students construct meaning and knowledge: they do not have meaning and knowledge handed to them in a book or lecture. Learning, then, is a process of students "making sense" of how things fit together; factual and procedural knowledge is built along the way.

## Linking Assessment with Instruction

A second, related assumption deals with the important role played by the context in which knowledge is constructed. For knowledge to be meaningful to students and for students to develop the attitude and manner of inquiry in a discipline, a culture of inquiry into the subject matter must be created (Lave, Smith, and Butler, 1988). The context is a culture of "doing" science or mathematics, not

---

Note: Research reported in this chapter was supported by grants from the National Science Foundation (No. SPA-8751511) and a University of California President's Grant for School Improvement. The views expressed in this chapter, however, are those of the authors and are not necessarily endorsed by either granting agency. This chapter is reprinted with permission from *Effective and Responsible Teaching: The New Synthesis,* edited by F.K. Oser, A. Dick, and J. Patry (San Francisco: Jossey-Bass, 1992). Copyright © 1992 by Jossey-Bass Inc., Publishers. All rights reserved.

receiving scientific or mathematical truths. Students learn not only a set of cognitive tools (for example, facts and procedures) but, more importantly, a set of beliefs about the subject and about valuing and caring for other students' contributions to knowledge construction. These cognitive and social tools, perpetuated in everyday practices and rituals of the culture, define the context in which knowledge is constructed.

One implication of these two assumptions for the development of alternative assessment technologies is that not just individually but also socially constructed solutions (for example, cooperative groups of students solving problems) should be encouraged (compare Resnick, 1988; Shavelson, Webb, Stasz, and McArthur, 1988). A second implication is that assessments should contain concrete, meaningful tasks. These tasks should respond to students' actions, providing feedback as they test ideas about problem solutions. A third implication is that tests should contain tasks for which there are alternative solutions. Hence, the tasks confronting students should be holistic in nature; the amount of time to solve them will exceed the usual thirty seconds allocated to an exercise provided by the teacher or a multiple-choice question on a test. Evaluation of performance on such tasks should capture the diversity of problem-solving strategies, compare them on a common metric, and provide credit for partial knowledge and well-reasoned, even if somewhat erroneous, solution strategies.

Changing the way we test, then, may well have a direct impact on what and how science and mathematics are taught. Teaching to tests that reward diverse problem solving in concrete, well-contextualized situations may directly influence teachers' instructional decisions to be consistent with the assumptions underlying effective and responsible teaching. In this chapter, we illustrate the nature of changes in teachers' decisions, goals, and plans through concrete examples and draw implications for educational reform.

## Alternative Technologies for Assessing Achievement

For the past several years, a team of researchers at the University of California, Santa Barbara, and the California University of Technology has been developing and evaluating alternative assessment technologies in science—assessments that are consistent with assumptions about effective and responsible teaching (Shavelson and others, 1991; Shavelson, Baxter, and Pine, forthcoming). These alternatives are based on students' performance of concrete, meaningful tasks (for example, a laboratory experiment to determine which of three paper towels soaks up the most water). Moreover, they are scored in a way designed to capture not just the "right answer" but also the reasonableness of the procedure used to carry out the task or solve the problem. Finally, these alternatives recognize the symmetry between testing and teaching; that is, a good assessment makes a good teaching activity, and a good teaching activity makes a good assessment. The project focused on upper elementary and middle school children (ages eleven through fourteen years) and provides a concrete example of the new vision of assessment and, indirectly, of effective and responsible teaching.

The researchers developed a number of alternative assessments, including observation of hands-on experimentation, notebooks based on the experiment, computer simulations of the experiment, and paper-and-pencil tests related to aspects of the experiment. We focus here on two examples. The first, the paper towels investigation, shows how assessments may reinforce science process understanding and encourage diversity of student performance. The second, an "electrical mysteries" simulation, reinforces application of knowledge in a problem-solving setting and demonstrates the symmetry of teaching and testing. Teaching to these assessments will change teachers' decisions about how to organize and present instruction.

For the paper towels investigation, students conducted an experiment to determine which of three kinds of paper towels soaked up the most water. A laboratory setup including towels, water, trays, graduated beakers, scissors, a scale, a timer, and so on was provided. Students were told that they could use all or some of the equipment available, as they felt necessary. After completing the investigation, students recorded procedures, findings, and conclusions in a lab notebook.

The diversity of student performances was striking. More than sixteen different procedural sequences were recorded—some legitimate, some flawed. The challenge to the researchers was to preserve this diversity and yet score the student performance on a common metric (see Baxter, Shavelson, Goldman, and Pine, 1992). To this end, a procedure-based scoring scheme was developed. It ordered the variety of performances as to their "scientific soundness" while preserving the actual procedures used. The scoring system recognized that several different procedures could result in the same quality of performance. For example, a student who poured equal amounts of water into three beakers, saturated a towel in each, removed the towels, and measured the amount of water remaining in each beaker would get the same score as a student who dipped each towel into a pitcher of water until saturated and then weighed each.

For the "electrical mysteries" simulation, students had to determine the contents of each of six mystery boxes (for example, battery and bulb) by connecting a circuit to the box. Students used a Macintosh computer and a mouse to connect the circuits, with the software emulating the behavior of an actual circuit. Students could connect a multitude of circuits on the screen at once if they so desired, or they could leave one completed circuit on the screen as an explicit model of what they thought was in the box and compare this circuit to the one hooked up to the box. Most germane to the present discussion was the finding that some students completed no fewer than 150 circuits in about thirty minutes. They had gone beyond the intended assessment to explore the nature of electrical circuits and to construct new knowledge through experimentation. The assessment had been viewed by the students as a teaching activity, not a test; hence the symmetry of teaching and testing.

This research on science assessments has several implications for learning how to teach science *effectively* if investigations such as the paper towels and electrical mysteries experiments are used to measure achievement. First, the materials typically found in elementary classrooms are going to have to change—from book learning to laboratory experiments, from pencil and paper to computers. Second, the typical arrangement of the classroom will have to change to accommodate the availability and use of lab equipment and computers. Third, teams of students are likely to be working together on a problem, partly because of limited access to equipment but more importantly because such an arrangement is consistent with the culture of doing science. Fourth, some concepts and norms of classroom discipline will need to change as students walk around and talk to one another. Fifth, writing up experiments, part of the science culture, will become routine. Indeed, science offers a concrete, meaningful occasion for practice in writing. And, sixth, teachers will have to deal with the diversity of solutions that students invent.

This research also has implications for what is considered *responsible* teaching. To prepare to take the test, students need to take responsibility for their own learning. The teacher becomes an orchestrator, responsible for developing learning activities and helping students construct knowledge. The goal is to create a culture in which students learn to work together cooperatively, to construct knowledge, and in the process to learn to value one another's unique contributions.

If teachers teach to the test, will they teach effectively and responsibly? To do so, we suspect that many teachers will have to change the way they plan and carry out instruction.

## Changes in Instructional Decision Making

To describe some of the changes in teaching that alternative assessment technologies may bring about, we have first to characterize instructional planning and action and the kinds of decisions that teachers make. Then we can characterize the impact of changes in the nature of assessments on these plans and actions.

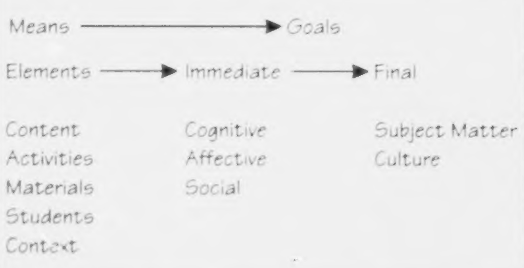
### A Framework for Teachers' Instructional Decisions

Research has shown that teachers' planning is opportunistic. That is, their focus is on creating a flow of activities that enact the curriculum while balancing multiple goals. The nature of those activities, the manner in which they are carried out, and students' acquisition of knowledge and values reflect the extent to which the teacher is effective and responsible.

Teachers' instructional planning and action focus on the instructional task (Doyle, 1986; Shavelson, 1983; Shavelson & Stern, 1981). As Figure 1 illustrates, this task contains two fundamental components: a set of goals and the means for reaching them. The immediate goals are not just cognitive but also intrapersonal (affective) and interpersonal (social skills and values). The final goals include subject-matter understanding and the creation of a subject-matter culture. This culture works together socially, values knowledge construction, and values the contributions of each and every student. The means include subject matter content, activities and materials, students' knowledge and motivation, and the classroom context (a sense of "groupness").

Teachers create instructional tasks when they plan instruction. The central concern is the construction of activities that connect goals with instructional means to create a flow of classroom instruction. Teachers' plans for instruction, which exert a strong influence on classroom teaching, are

FIGURE 1
COMPONENTS OF THE INSTRUCTIONAL



BEST COPY AVAILABLE

guided by a conceptual framework or schema (Shavelson, 1986). An instructional schema includes slots for instructional means (content, activities, etc.) that are linked to goals. When teachers plan, they fill the slots in their instructional schema with content, activities, etc..

An instructional schema contains three kinds of slots. One kind of slot contains *subject-matter content*. These slots hold concepts and skills to be taught, and are linked together to preserve the relations among the concepts and skills. A second kind of slot contains activities students will engage in. These slots are linked together in a time sequence forming a progression of activities or an instructional "*script*" for the lesson. This script, played out in the classroom, is predictable to students and teacher alike because it has been played out, albeit with somewhat different content and activities, from one lesson to the next. The third kind of slot contains information for creating a physical environment or "*scene*" for the classroom in which the script will be played out. Scenes (e.g., reading circle, seat work), familiar to both teacher and student, facilitate the enactment of the activities (e.g., Berliner, 1983).

Taken together scripts and scenes create the sociocultural context of the classroom. They are greater than the sum of their parts, creating a whole culture—a manner of approaching subject matter and the individual and social construction of knowledge.

### Impact of Alternative Technologies on Instructional Decisions

As we have seen, the alternative assessment technologies are based on two important assumptions about the nature of learning: that knowledge is constructed and that this construction takes place within a cultural context of a discipline. If teachers want to teach to the new tests, they will have to align their professional knowledge—schemata—with these assumptions.

The teacher's primary role—conveyor of knowledge—must change. The teacher must provide opportunities for students to make meaning and construct new knowledge. To do so, the teacher needs to value the capabilities and contributions of all students and establish a culture of doing mathematics or science. This culture is one in which its members—teacher and students together—find problems, debate and establish criteria for solving them, and construct alternative solutions.

To create this culture, the teacher must possess "a strong mastery of the practice of mathematics" and science and have "the autonomy to engage with learners in the practice of problem solving in inventive ways" (Lave, Smith, and Butler, 1988, p. 77). Moreover, the teacher has the responsibility of modeling care and respect for student contributions, creating a strong sense of self and capability in the learners. In short, the usual goals of instruction (often acquisition of facts and basic skills) and the usual scripts and scenes will have to be replaced by new ones. They may shift from primarily cognitive—teaching basic operations in mathematics or definitions in science—to cultural—constructing a supportive culture in a subject matter. Within this culture, teachers will seek to intertwine intellect, values, attitudes, and social interaction.

To teach to the alternative assessments, teachers will need a firm grounding in subject matter knowledge. They may need a much greater understanding of mathematics, for example, than they now possess if they are to facilitate students' recognition of a mathematical problem and its representation in alternative symbolic forms, identification of criteria against which to evaluate solutions, and so on (Lave, Smith, and Butler, 1988; Leinhardt and Smith, 1985). The usual pedagogical assumptions and rationales that underlie teachers' instructional decisions, such as those that lead to whole class instruction with the teacher as primary information conveyor ("direct instruction"), may also be called into question; assumptions and values embodied in the constructivist reform will affect teaching scripts and scenes and lead to different instructional decisions and practices.

The most common teaching *script* in the United States is recitation: teachers ask questions and students give quick, short answers; teachers talk about 80 percent of the time and students 20 percent. A second common script is seatwork: after a brief explanation of a concept in mathematics or science, students work individually on exercises or definitions in a workbook (Romberg and Carpenter, 1986). Neither script gives rise to an intellectual community respecting individual and joint contributions to knowledge or to a culture within a subject matter.

Teachers will need to acquire new scripts, scripts that are just emerging. Their task will be to facilitate the development of an intellectual community. A new set of responsibilities and values regarding knowledge and its construction may be needed. No longer will teachers make decisions within the constraints of a fifty-minute hour. Knowledge construction does not recognize these time barriers. Hands-on activities and apparatus will find their way into the script. Sequencing will be guided by student exploration, unpredictable in the specifics but predictable at the macro level. The pace will be set by the students, as will the timing of activities. The content will be built from the concrete activities. In short, teachers' decisions about the content, materials, sequencing, pacing, and timing of classroom activities will change.

The physical and social organization and operation of classrooms may very well change substantially as well. Typical scenes associated with seatwork and recitation, with no social contact, will decrease in frequency as task-oriented social interaction and knowledge construction emerge. Students will no longer be seated at individual desks; larger work spaces will be needed. Decisions about allocation of classroom space, equipment, and materials will have to serve the need for students to interact and to use equipment in the discovery of knowledge.

For most teachers, this view of instruction and the decisions that give rise to it are not routine today. The nature of the scripts and scenes that teachers instantiate to motivate and support knowledge construction are not well established (but see, for example, Lampert, Chapter Nineteen in this volume). They will emerge and become recognizable if the alternative assessments have the impact expected. They will become a necessary precursor of teaching to the test. These new teaching routines are going to challenge the boundaries of schooling as we know them today. They may very well lead to the restructuring of schooling. However, a caveat is in order. Without substantial commitment to training and supervision of teachers, without public support, without the necessary materials and facilities, there are no guarantees that the picture of effective and responsible teaching painted here will emerge in the curriculum reform. Indeed, even with this support, the reform goals may get warped.

An anecdote may serve to illustrate the importance of our concern. One of the teachers whose classes participated in our research knew of the paper towels experiment. Teaching to the test, she instructed her students to saturate the towels completely, using the timer to ensure that at least ten minutes were given to saturating each towel separately (a total of thirty minutes for saturation!), when, in reality, the towels could be saturated in a matter of seconds. The result was that the students performed in a clearly stylized manner. The teacher had informed them, perhaps unintentionally, that science is a set of precise steps that must be carried out invariably, regardless of whether they make sense. This was not particularly conducive to scientific exploration; although the teacher could teach the students how to "do" the experiment, what was missing was the cultural context of doing science.

## Implications: A Basis for Restructuring Education

One of the most often cited solutions to educational problems in the United States is "restructuring." Shake up the schools organizationally, the assumption goes, and when the pieces are reassembled, things will be better. For example, place greater decision-making responsibility with teachers and less with administrators: shifting decision making to those more closely linked to the function of schooling—teaching and learning—will (somehow) improve education. Or place decision making at the local school site, not at a central office. The incentives for quality schools, politicians argue, reside in the community. Somehow, nonprofessionals will discover a way to improve education.

None of these "solutions" paints a picture of what effective and responsible teaching would look like and why student learning and performance would improve. These suggestions place form—organization of schools before function. Suppose that the function of schools were to create an intellectual community in which students and teachers value one another and together construct knowledge and use it in meaningful ways to solve problems, resolve dilemmas, or make meaning out of the unexpected. Suppose further that our accountability systems held teachers accountable with assessments congruent with school functions—not just in individual academic achievement but also in the ability to work together constructing knowledge and solving problems. Teachers would design instruction to this end. The design decisions would lead to changes in scripts and classroom scenes. The physical organization of the classroom, the temporal organization of the school day, and the relations among teachers and students would reflect the functions of education. If teachers, educational systems, policy makers, and the public respond to and support changes in assessment technologies as we envision, changes in teaching may very well lead to restructuring schools in an effective and responsible way.

## REFERENCES

Anderson, R. "Some Reflections on the Acquisition of Knowledge." *Educational Researcher,* 1984, *13* (9), 5-10.

Baxter, G. P., Shavelson, R. J., Goldman, S. R., and Pine, J. "Evaluation of Procedure-Based Scoring for Hands-On Science Assessment." *Journal of Educational Measurement,* 1992, *29* (1), 1-17.

Berliner, D. C. "Developing Conceptions of Classroom Environments: Some Light on the T in Classroom Studies of ATI." *Educational Psychologist,* 1983, *18* (1), 1-13.

Doyle, W. "Classroom Organization and Management." In M. C. Wittrock (ed.), *Handbook of Research on Teaching.* (3rd ed.) New York: Macmillan, 1986.

Glaser, R. "Education and Thinking: The Role of Knowledge." *American Psychologist, 1984, 39* (2), 93-104.

Gough, P. B. (ed.). Special section on testing. *Phi Delta Kappan,* 1989, *70* (9), 683-725.

Lave, J., Smith, S., and Butler, M. "Problem Solving as Everyday Practice." In R. I. Charles and E. A. Silver (eds.), *The Teaching and Assessing of Mathematical Problem Solving.* Hillsdale, N.J.: Erlbaum, 1988.

Leinhardt, G., and Smith, D. A. "Expertise in Mathematics Instruction: Subject Matter Knowledge." *Journal of Educational Psychology,* 1985, *77,* (3), 247-271.

Mulls, I. V., and Jenkins, L. B. *The Science Report Card: Elements of Risk and Recovery.* National Assessment of Educational Progress report no. 17-S-01. Princeton, N.J.: Educational Testing Service, 1988.

National Commission on Excellence in Education. *A Nation at Risk: The Imperative for Educational Reform.* Washington, D.C.: U.S. Government Printing Office, 1983.

Nickerson, R. S. (ed.). "New Directions in Educational Assessment." *Educational Researcher,* 1989, *18* (9), 3-32.

Resnick, L. B. *Education and Learning to Think.* Washington, D.C.: National Academy Press, 1987.

Resnick, L. B. "Treating Mathematics as an Ill-Structured Discipline." In R. I. Charles and E. A. Silver (eds.), *The Teaching and Assessing of Mathematical Problem Solving.* Hillsdale, N.J.: Erlbaum, 1988.

Romberg, T. A., and Carpenter, T. P. "Research on Teaching and Learning Mathematics: Two Disciplines of Scientific Inquiry." In M. C. Wittrock (ed.), *Handbook of Research on Teaching.* (3rd ed.) New York: Macmillan, 1986.

Shanker, A. "The End of the Traditional Model of Schooling—and a Proposal for Using Incentives to Restructure Our Public Schools." *Phi Delta Kappan,* 1990, *71* (5), 345-357.

Shavelson, R. J. "Review of Research on Teachers' Pedagogical Judgments, Plans, and Decisions." *Elementary School Journal,* 1983, *83* (4) 392-413.

Shavelson, R. J. *Pensamientos de los Profesor y Toma de Decisiones* [Teacher cognition and decision making]. Seville, Spain: University of Seville, 1986.

Shavelson, R. J., Baxter, G. P., and Pine, J. "Performance Assessments in Science." *Applied Measurement in Education,* forthcoming.

Shavelson, R. J., Carey, N. B., and Webb, N. M. "Achievement Indicators: Options for a Powerful Policy Instrument." *Phi Delta Kappan,* 1990, *71* (9), 692-697.

Shavelson, R. J., and Stern, P. "Research on Teachers' Pedagogical Thoughts, Judgments, Decisions, and Behavior." *Review of Educational Research,* 1981, *51* (4), 455-498.

Shavelson, R. J., Webb, N. M., Stasz, C., and McArthur, D. "Teaching Mathematical Problem Solving: Insights from Teachers and Tutors." In R. I. Charles and E. A. Silver (eds.), *The Teaching and Assessing of Mathematical Problem Solving.* Hillsdale, N.J.: Erlbaum, 1988.

Shavelson, R. J., and others. "Alternative Technologies for Large-Scale Assessment: Instrument of Education Reform." *School Effectiveness and School Improvement,* 1991, *2,* 1-18.

Smith, M. L. "Put to the Test: The Effects of External Testing on Teachers." *Educational Researcher,* 1991, *20* (5), 8-11.

Walker, D. F., and Schaffarzick, J. "Comparing Curricula." *Review of Educational Research,* 1974, *44,* (1), 83-111.

Wiggins, G. "A Trust Test: Toward More Authentic and Equitable Assessment." *Phi Delta Kappan,* 1989, *70* (9), 703-713.

# IV-8

# What You Assess May *Not* Be What You Get

Thomas R. Guskey
University of Kentucky

Few innovations in education have caught on as quickly as performance-based assessment. Nearly every current reform initiative includes a provision to assess students' performance on complex learning tasks.

The tasks specified include essays, demonstrations, computer simulations, performance events, portfolios of students' work, and open-ended questions and problems. Collectively, these measures are referred to as *authentic* assessments because they are valuable activities in themselves and involve the performance of tasks that are directly related to real-world problems (Linn et al. 1991).

Two major factors have intensified the interest in performance-based assessment. First, advances in cognitive science (see, for example, Resnick 1985 and 1987, or Shuell 1986) have compelled educators to acknowledge how complex learning is and how diverse are the means needed to assess learning fully and fairly.

Second, many educators have recognized the limitations of assessment systems that have relied on multiple-choice, standardized achievement tests. Researchers have found that such systems, especially those used to ensure accountability, encourage teachers to skew their instruction to the basic skills assessed in the tests (Haladyna et al. 1991, Shephard 1990). As a result, the curriculum narrows and the validity of information gathered from the tests diminishes (Mehrens and Kaminski 1989, Shephard 1989). These two effects are commonly amplified in schools serving at-risk and disadvantaged students because such schools are under great pressure to show improvement in test scores (Herman 1992).

Advocates of new performance-based assessments believe that if teachers are going to teach to tests, the tests (or other forms of assessment) should be worth teaching to. Then, reformers hope, first-rate tests will call forth first-rate instruction. For example, assessment devices that tap higher-order thinking skills will elicit instructional practices that emphasize and develop higher-order thinking skills. An added benefit is that the performance-based assessments are likely to become an integral part of the instructional process, rather than a separate, after-the-fact check on student learning (Wiggins 1989a). The distinction between instruction and assessment would thus become "seamless."

Some educators have carried this vision a step farther, suggesting that authentic, performance-based assessments could actually *drive* instructional improvements (McLaughlin 1991). This approach is called measurement-driven instruction, or MDI. (Popham 1987, Popham et al. 1985).[1]

---

First published in *Educational Leadership* 51, 6: 51–54.

## Reform in Kentucky

Kentucky recently enacted reform legislation that takes the measurement-driven instruction approach. The new law, the Kentucky Education Reform Act (KERA), is one of the most comprehensive pieces of educational reform legislation ever enacted in the United States. It addresses administration, governance and finance, school organization, professional development, curriculum, assessment, and accountability.

A significant component of KERA is a student assessment program called the Kentucky Instructional Results Information System. KIRIS is a multifaceted program that evaluates:

- portfolios of students' work in writing and mathematics;
- students' achievement on "performance events" in the areas of mathematics, science, social studies, arts and humanities, and vocational education/practical living; and
- students' scores on "transitional tests," which include both multiple-choice and open-ended items similar to those in National Assessment of Educational Progress tests. Transitional tests assess performance in reading, writing, mathematics, science, social studies, arts and humanities, and vocational education/practical living.

KIRIS is a high-stakes assessment program. That means that results from the assessments will be used to grant financial rewards to schools that improve significantly and to levy sanctions against schools that fail to show progress (Foster 1991). The high-stakes nature of the assessment program is what makes KERA a measurement-driven reform effort (Guskey 1994).

KERA and KIRIS are particularly interesting to educators and policymakers for two important reasons. First, although KERA is not the first reform effort to include a comprehensive assessment system, it is the first to be driven by assessments that are *primarily performance-based.* Second, KERA is the first statewide reform effort with *high-stakes* performance-based assessments.

High-stakes assessment itself is not new to Kentucky educators. During the 1980s, results from the administration of a statewide test known as the Kentucky Essential Skills Test (KEST) were used to rank school districts throughout the Commonwealth and to dispense rewards and sanctions (Guskey and Kifer 1990). KIRIS is a sharp departure from KEST, however, in that KEST was composed entirely of multiple-choice items designed to assess basic skills.

## The Vitali Study

The implementation of these two conceptually different high-stakes statewide assessment programs, both used within a 10-year period, presented an excellent opportunity to compare the impact of each program. One of my doctoral students, Gary Vitali, recently set out to determine the impact of such assessment systems on teachers' instructional practices.

Vitali's study involved extensive teacher interviews, several teacher questionnaires, and classroom observations (Vitali 1993). The findings offer new insights into the complexities of measurement-driven reform and also challenge the notion that what you test is invariably what you get in the classroom.

Vitali's findings support those of other researchers who have found that multiple-choice, standardized achievement measures employed for accountability purposes *do* focus instruction on the content of the tests. Most teachers narrow the curriculum and focus instruction on basic skills. They do so, Vitali found, because they want their students to do well on the tests, whether or not the teachers believe that the content and skills being measured are important. He also discovered that most teachers think teaching to standardized tests is fairly easy to do. After all, narrowing instruction is easier than broadening it, and most teachers reported that their instructional materials are generally aligned with a basic-skills orientation.

The performance-based assessment program, on the other hand, resulted in only modest changes in teachers' instructional practices. A few teachers who recognized that most of the performance tasks and portfolio entries required students to do some writing did respond by incorporating additional writing activities in their daily lessons and in classroom assessments. For the vast majority of teachers, though, lesson plans, classroom activities, and evaluations of student learning remained unchanged.

This finding was all the more surprising in view of teachers' positive attitudes toward performance-based assessments. Teachers regarded these more broadly based assessments as better measures of student learning than multiple-choice, standardized achievement tests.

In interviews and questionnaires, Vitali sought to determine why teachers did not make more significant adaptations, especially considering the high-stakes nature of the Kentucky assessment program. He discovered that, simply put, teachers did not know how to teach to the performance-based assessments, nor did they believe that they could do so within their current time constraints.

Although most teachers said they felt "under the gun" to adapt instructional practices to the performance-based assessments, the transition seemed insurmountable because it required professional training and time the teachers did not have. The need for training seemed especially critical since the realignment would involve an expansion both of what is taught (curriculum) and how it is taught (methods).

The teachers' perceptions were borne out. Although Vitali found the vast majority of teachers to be dedicated, hard-working individuals who want their students to do well, he also discovered that, in general, teachers were ill-prepared to adapt their instructional practices to the new demands of a more authentic, performance-based assessment program. Most teachers had scant knowledge, personal background, experiences, or formal training with the various types of performance-based assessments or ways to use them as instructional tools. The only training that most teachers had received was scattered, one-day staff development workshops.

The lack of personal experience and professional training in instructional techniques that might help students prepare for performance-based assessments was a widespread problem that seemed to affect both elementary and secondary teachers. Many respondents also stressed that the lack of appropriate teaching materials was a problem.

Teachers perceived two general types of time pressures. First, teachers reported that they were being required to do more and teach more, without any increase in the amount of time allowed for planning or instruction. (Secondary teachers mentioned this obstacle more often, possibly because they have traditionally been more content-oriented and thus see performance-based assessments as a more drastic change than do skills-oriented elementary teachers.) Second, most teachers believed that performance-based assessments would require a lot more time to administer and score.

These perceptions of little time and lots of extra work, combined with inadequate experience, training, and materials, appeared to keep most teachers frozen in virtually the same instructional patterns that they had before the new assessment system. Thus Vitali concluded that "what you test *may not* be what you get" when performance-based assessments are the primary testing tool and teachers have neither adequate time nor sufficient training to teach to the test.

## Accountability Won't Ensure Success

Although the Vitali study has limitations and its findings will require confirmation and elaboration by other researchers, the results clearly indicate that instituting a high-stakes, performance-based assessment program, even one as thoughtfully designed and as carefully implemented as Kentucky's KIRIS program, is not enough to bring about significant change in the instructional practices of most teachers. Adapting instructional practices to performance-based assessments, the study shows, is a much more complex process than many advocates of measurement-driven instruction assume. Bridging the chasm between authentic assessment and authentic classroom practice will require well-designed assessments, but it will also demand a substantial amount of additional time, resources, and training opportunities.

If a performance-based assessment program is to evoke more stimulating, intellectually challenging tasks for students, extensive professional development opportunities for teachers will need to accompany the assessment program. These opportunities could offer ideas on how to design activities that promote authentic learning, suggest instructional materials that involve students in high-level processes, and recommend classroom assessment designs that are more performance-based (Stiggins 1987). Adequate treatment of these topics will certainly require more extensive time commitments than a one-day inservice program. Further, because the challenge involves the expansion of teachers' expertise and instructional repertoires, regular follow-up and continuous support will also be important factors (Guskey 1991).

Thus the lesson from Vitali's study is clear. Performance-based assessments, by themselves, appear to be insufficient to bring about significant change in the instructional practices of most classroom teachers, and without change in instructional practice, improvement in student learning cannot be expected. On the other hand, combining authentic, performance-based assessments with high-quality professional development opportunities to help teachers align instruction with improved assessments will make significant advances in student learning much more likely.

[1] Some critics, like Airasian (1988) and Worthen (1993), have expressed reservations about MDI, and Cizek (1991, 1993) and Mathison (1990) are among those who view the approach as unethical. Even so, proponents of performance-based assessments argue that they are more likely to engage students in complex intellectual challenges than are the uninspired teaching practices—so common today—that promote only memorization of unrelated bits of information (Wiggins 1989b).

## REFERENCES

Airasian, P. W. (1988). "Measurement-Driven Instruction: A Closer Look." *Educational Measurement: Issues and Practice* 7, 4: 6-11.

Cizek, G. J. (1991). "Innovation or Enervation?" *Phi Delta Kappan* 72: 695-699.

Cizek, G. J. (1993). "Rethinking Psychometricians' Beliefs about Learning." *Educational Researcher* 22, 4: 4-9.

Foster, J. D. (1991). "The Role of Accountability in Kentucky's Education Reform Act of 1990." *Educational Leadership* 48, 5: 34-36.

Guskey, T. R. (1991). "Enhancing the Effectiveness of Professional Development Programs." *Journal of Educational and Psychological Consultation* 2: 239-247.

Guskey, T. R., ed. (1994). *High Stakes Performance Assessment: Perspectives on Kentucky's Educational Reform.* Newbury Park, Calif.: Corwin Press.

Guskey, T. R., and E. W. Kifer. (1990). "Ranking School Districts on the Basis of Statewide Test Results. Is it Meaningful or Misleading?" *Educational Measurement: Issues and Practice* 9, 1: 11-16.

Haladyna, T. M., S. B. Nolen, and N. S. Haas. (1991). "Raising Standardized Achievement Test Scores and the Origins of Test Score Pollution." *Educational Researcher* 20, 5: 2-7.

Herman, J. L. (1992). "What Research Tells Us about Good Assessment." *Educational Leadership* 49, 8: 74-78.

Linn, R. L., E. L. Baker, and S. B. Dunbar. (1991). "Complex, Performance-Based Assessment: Expectations and Validation Criteria." *Educational Researcher* 20, 8: 15-21.

Mathison, S. (1990). "Controlling Curricular Change Through State-Mandated Testing: Ethical Issues." Paper presented at the annual meeting of the American Educational Research Association, Boston, Mass.

McLaughlin, M. W. (1991). "Test-Based Accountability as a Reform Strategy." *Phi Delta Kappan* 73: 248-251.

Mehrens, W. A., and J. Kaminski. (1989). "Methods of Improving Standardized Test Scores: Fruitful, Fruitless, or Fraudulent?" *Educational Measurement: Issues and Practice* 8, 1: 14-22.

Popham, W. J. (1987). "The Merits of Measurement-Driven Instruction." *Phi Delta Kappan* 68: 679-682.

Popham, W. J., K. L. Cruse, S. C. Rankin, P. D. Sandifer, and P. L. Williams. (1985). "Measurement-Driven Instruction: It's on the Road." *Phi Delta Kappan* 66: 628-634.

Resnick, L. B. (1985). "Cognition and Instruction: Recent Theories of Human Competence." In *Master Lecture Series: Vol. 4, Psychology and Learning*, edited by B. L. Hammonds. Washington, D.C.: American Psychological Association.

Resnick, L. B. (1987). "Constructing Knowledge in School." In *Development and Learning: Conflict or Congruence?*, edited by L. S. Liben. Hillsdale, N.J.: Erlbaum.

Shephard, L. A. (1989). "Why We Need Better Assessments." *Educational Leadership* 46, 7: 4-9.

Shephard, L. A. (1990). "Inflated Test Score Gains: Is the Problem Old Norms or Teaching to the Test?" *Educational Measurement: Issues and Practice* 9, 3: 15-22.

Shuell, T. J. (1986). "Cognitive Conceptions of Learning." *Review of Educational Research* 56: 411-436.

Stiggins, R. (1987). "Design and Development of Performance Assessments." *Educational Measurement: Issues and Practice* 6, 3: 33-42.

Vitali, G. J. (1993). "Factors Influencing Teachers' Assessment and Instructional Practices in an Assessment-Driven Educational Reform." Doctoral diss., University of Kentucky.

Wiggins, G. (1989a). "Teaching to the (Authentic) Test." *Educational Leadership* 46, 7: 41-47.

Wiggins, G. (1989b). "A True Test: Toward More Authentic and Equitable Assessment." *Phi Delta Kappan* 70: 703-713.

Worthen, B. R. (1993). "Critical Issues That Will Determine the Future of Alternative Assessment." *Phi Delta Kappan*: 444-456.

# IV-9

# Evaluating the Teaching and Learning of Literacy

FRANCINE Z. STAYTER
SUNY, ONEONTA

PETER JOHNSTON
SUNY, ALBANY

Reprinted with permission from *Reading and Writing Together: New Perspectives for the Classroom*, edited by Timothy Shanahan (Norwood, Mass.: Christopher-Gordon Publishers, Inc.), pp. 253–271. Copyright © 1990 by Christopher-Gordon Publishers, Inc. All rights reserved.

*In this thoughtful chapter, Stayter and Johnston propose a revitalization of teacher thinking and student self-awareness through assessment and evaluation. They argue that it is not enough to put reading and writing together but that we need to integrate instruction and assessment as well. They suggest many penetrating ways of knowing that are directly appropriate to the teaching recommendations offered throughout the previous chapters by the other authors, and their emphasis on teacher thinking and student involvement bring this volume full circle, amplifying the themes introduced in my first chapter.*

# Evaluating the Teaching and Learning of Literacy

**Francine Z. Stayter and Peter Johnston**

> Those who believe in our ability do more than stimulate us. They create for us an atmosphere in which it becomes easier for us to succeed.
>
> John Lancaster Spaulding

The research of recent years has begun to heal the splits that exist between reading, writing, teaching, and learning, as well as those that exist between research, theory, and practice. There has been an increasing realization that all represent acts of composition and reflection. Evaluation, on the other hand, has remained separate from teaching, learning, reading, and writing (Johnston, 1987). When people think of the issue of evaluation, the most frequent association that comes to mind is large-scale standardized testing in which student populations can be compared with one another and teacher populations can be compared with one another across time and contexts. Yet of all the evaluations, it is the day-to-day, moment-to-moment transactions between teachers and students and students among themselves that have the greater influence on what students learn about being literate and what teachers learn about students and their own teaching practice. Evaluation is central to the effective teaching and learning of literacy and the way we evaluate literacy instruction affects, and it is affected by both the means and the ends of literacy teaching and learning.

The ultimate purpose of evaluation is to learn about and improve teaching and learning (Johnston, 1987). Indeed, instead of talking about "evaluation," we might equally correctly talk about "learning." But what we learn is very much influenced by what we choose to look at, how we frame what we look at, and the context in which we work. If we base our instructional decisions principally on standardized test scores or criterion-referenced basal reader tests we are likely to make different instructional decisions than if we base our decisions principally on interviews, observations, and demonstrations. In turn, these decisions will influence the way we interact with students and the way they come to know literacy and themselves as literate individuals. The same is true of the evaluation of teaching. A public (or an administration) that knows teachers principally through student results on standardized tests will make different decisions than one that knows its teachers through involvement. These perspectives will be reflected in the ways teachers view themselves, their students, and their practice. Evaluation is, whether we like to admit it or not, a fundamentally political activity; it can be a reflective act that individuals and communities engage in to improve teaching and learning or a tool for controlling who gets to do what. We would like to see greater recognition of the value-laden nature of evaluation as its Latin root "value" suggests.

Currently, evaluation is separated from instruction and based on non-involvement. It is not seen as a constructive, value-laden, and reflective activity. Rather, it is seen as an objective, value-free, non-reactive activity. We would like to see evaluation back in its rightful place, along with reading, writing, teaching, and learning, as an involving, reflective process in which one learns about and develops one's practice (of teaching or being literate) and one's self as a practitioner. We believe that as teachers we seek to help not some generic construct called "students" but rather real learners, people with names, who have been active meaning makers long before they reached school. Thus, in this chapter we will focus on the nature of classroom evaluation and its consequences. We will assert the importance of the common or garden variety of evaluative transactions that often go unnoticed by educators but which have a powerful impact on students' and teachers' ways of viewing themselves and their literate activity. We will assert that whether we are evaluating in formal, standardized formats or in the day-to-day transactions with our students, all evaluation is ultimately a social transaction in which we give as much information as we get. Finally, we assume that evaluation is constructive, that we make sense out of our students' and our own behavior, and that different evaluators will construct a different meaning from the same set of observations.

## What We See and How We Talk About It.

Although evaluation has been kept separate, it has a lot in common with reading, writing, and teaching. It, too, is constructive. Teachers read their students and their classrooms. Just as readers' knowledge of the world and values are reflected in the meaning they construct from reading a text (Rosenblatt, 1984), what teachers know and how they know it will be reflected in their evaluations of their students and themselves. A teacher's knowledge affects not only what is evaluated, but how a teacher will use the evaluation to inform instruction. In a recent study we asked teachers to describe some of their students' literacy development—a kind of oral free recall of their reading of the students' development. In general, if teachers knew a lot about children's literature, this knowledge was reflected in their descriptions of children's literacy development. If teachers knew little about children's literature, that too was reflected. For example, compare the following recalls by two teachers whose knowledge of literature differed substantially.

> So he would stick to those kinds of books . . . *Three Bears, The Business Letter*, those kinds of books. Then gradually he branched out a little bit to . . . *Fortunately* is one of Mercer Mayer's books, and the Clifford books, and that type of thing . . . . . . And now, I guess he'd have to be considered one of the Patricia Riley Giff experts . . . those books are not real easy for him, but he just likes the character so much. I think he's in love with Ms. Rooney, with her classroom that she has. That he's willing to spend that extra time getting to know more about those children in that room, in that book, and he really tries very hard to be very self-sufficient reading those books of hers . . . He's tried the mystery ones that she's come out with . . . [and that was only one quarter of the description of a single child]

Compare that with a description by another teacher whose knowledge of literature was more limited:

> From the beginning of him coming into my room he could not read. He's eleven years old, in third grade, and very little sight vocabulary. Now he's able to read, himself. He's made progress. He'll still be placed in special ed in a couple of weeks, but I do see progress. [This was the entire description]

These descriptions reflected several differences between the teachers' readings of their students. First, there is the sheer quantity of knowledge the teachers had gathered about their students' development. Each teacher was asked to describe a student they knew well. Second, there is the extent to which the description was tied to children's literature. Third, there were differences in the personal nature of the description. Some described them in terms of their involvement with particular books, others described them in terms of "placement levels" ("he's a 3.2"). These teachers read their students differently and certainly respond to their perceived instructional needs differently. Just as written language is open to legitimately different interpretations by different readers, so too can children be legitimately read differently by different teachers. Because individuals differ in their use of language and in their interpersonal transactions, different teachers will compose their responses to students and their evaluations to other interested parties differently. Though diversity in interpretation and in language use should be celebrated, we must be aware that every reading and every evaluation has its consequences.

The techniques we use to evaluate, the focus of the evaluation, and the language we use will all have an impact. Until recently our concern was primarily: "What techniques or procedures can a teacher use to learn about her students' development?" Simply selecting the techniques was a difficult problem for many; once selected, however, the teachers administered, the students took, and that was it. Testing *was* the evaluative solution. More recently we have come to realize that not only must we ask: "From what techniques or procedures will we learn the most in order to support continued learning?" but we must also ask: "What will our use of these techniques tell our students about reading, writing, learning, and being literate?" What we choose to evaluate and how we choose to evaluate delivers powerful messages to students about those things we value. Thus, we need to be concerned with how our evaluations impact on how students will ultimately evaluate themselves and their learning. An example of this impact can be seen in the recollections of a seventh grade student we interviewed. He was new to the school and recounted being in remediation since fourth grade in his other school. In reflecting on his development as a reader, the most cogent recollection was that for all of grades four, five, and six he was in the same book. The book, he recalled, had a circle on the back with pie-shaped pieces. The book he used had four wedges, which for two years did not add up to his grade, much to his shame. He recalled being able to read the text but being unable to pass the magazine test that would allow him to go on. In his case the evaluation provided for by the text did not seem to inform instruction, as the student was simply recycled each year.

Students view their learning and their sense of worth through the lens we help them construct unless they cannot bear to look through it. If the evaluative responses students receive are generally positive, that is how students will tend to view their efforts. If the evaluative responses are negative—"Stupid mistakes!" or frowning face stickers—that, too, becomes a way students begin to view their efforts. What teachers say to students conveys a powerful message to students of what is of value in this classroom. If we consistently comment to students about their error-free oral reading or the correctness of the spelling on a draft, they are likely to learn that what counts is error-free behavior in *all* literacy events. There is, of course, a place for presentational forms of literacy when students have had the time and option to prepare for performance (even prodigies practice), but our consistent emphasis on error-free, first (which is also final) draft behavior constrains student learning, and teacher learning, as well. Bruner (1986) argues that very little new learning occurs without error. Other researchers (Bissex 1980, Clay 1985, Goodman, et. al. 1987) indicate that errors give us an indication of how the student understands the particular literate activity. This is especially useful evaluative information that can be used to inform optimal instruction. When we eradicate error, when we make students feel the need to offer only the safe and correct, we limit their opportunities to learn and our own opportunities to learn about students' developing understandings. If we, instead, treat errors as opportunities to learn, students are more likely to explore new possibilities and more likely to deal positively with their errors. Again, the same is true of teachers and evaluations of their teaching. When errors carry heavy penalties, we can expect little of an imaginative nature going on in the classroom, and we can expect a good deal of denial of responsibility for errors and refusal to learn from them.

For many years classroom evaluation has focused on what our students could not do. Much of that grew out of packaged assessments that knew neither our classrooms nor our students. Student response on assessment tasks was viewed as right or wrong. Teachers found what was wrong, taught to it, and reassessed for evidence of mastery. More recently we have begun to look at what is right and what students are working on (Wertsch, 1979; Applebee & Langer, 1983). For example, in one ongoing study, a first grade teacher talked about helping her first grade writers. She noticed that many students used *z* when they wanted to turn singular nouns into plurals. In class one day she told them that she had noticed that many of them were doing this. "That's really thinking," she said to them. "That makes a lot of sense." Then, together, they discovered how it was done in their commercial big book. Together they repeated some plural words to hear and feel the final sound in the words. She let them know that they were good thinkers to use *z*. "It really does sound that way, doesn't it?" In the days that followed she noticed that many of her students felt very proud that they could now do something they hadn't been able to do before. One student read through all of his journal and, on his own, carefully crossed out all the *z*'s, putting *s*'s in their place. He told his mother that he was becoming a book writer because now he knew how to do something like book writers. He was not discouraged by what he previously could not do; he was motivated by what he could now do and wanted to do more. This same observation could have been read and responded to negatively just as easily: "When I read your journals, I saw that you were spelling words like cars and dogs wrong. This is the right way to write it." Both teachers observed non-standard spelling. The first teacher saw a pattern of learning and used instruction to build on it. The second teacher saw error and eradicated it without seeing what the students were trying to do. Consistently focusing on the positive helps children view their work in a similar positive light.

In the same way, the frequency and primacy of our responses to particular aspects of reading or writing carries a message. Consistently focusing on the content of a piece of writing rather than the convention conveys the idea that the content is most important. But it does more than that. When a teacher, and members of the learning community who model their responses on those of the teacher, consistently focus on the content of a student's writing before the conventions of it, the student learns that someone is interested in what they have to say and that what they have to say is worth saying. That, in writing, is the most important part. A consequence is a motivation to get the message right, as a concern for that audience, and to persist with the writing. Consistent primary focus on the conventions of writing will generally produce conventionally accurate but voiceless and uninteresting writing for which students show no ownership. For example, we observed one first grade writer who spent the first eight weeks of school limiting herself to the few words she had been taught. Her writing was limited to "I love" and "I love Mommy." When she realized that what she had to say was more important than the spelling of the words, she began writing more extended discourse. Typical of her writing were the position pieces she wrote; an analytic form of expression. In these she would state position and then give evidence for that position. For example:

> It is snowing.
> I like snow.
> I like to ski.
> I like to sled when it snows.

Writing in a community of writers provides opportunities for students to provide evaluative feedback for each other. The pattern of evaluative responses modeled by the teacher is generally replicated in the student community. In such a community, when children read their writing to others, their audience hears and responds primarily to the message of the piece. A common response format includes:

- — what you heard,
- — a part you particularly liked, and
- — any questions or confusions you have

For students to respond within this format requires careful listening. When writers find that they are being heard, they begin to find their voice. They are more able to write pieces that audiences find interesting because they receive positive, focused feedback. In time this structure becomes an internal one that students use to evaluate their own work as well as the work of others (Applebee & Langer, 1983). A middle-school student in a study that looked at the development of literate thinking showed this internalization, "I know the members of my group really like a lot of detail in my writing. They say they can 'see it' when I do that." Students in such contexts do not necessarily eschew conventions of spelling and punctuation, but they work first on the meaning of the piece and edit when a text is to be published. Knowing you have something worth saying is strong motivation to make the text's presentational form as correct as possible.

## The Context in Which We Learn

The context in which we learn has a considerable impact on the kinds of transactions that take place in the classroom, the frequency of the transactions, and the way they are understood. For example, some teachers or their administrators value silence in the classroom, believing that if children are talking they are not learning. This effectively cuts off the development of community, as well as communication. But community can be reduced in other ways, too. For example, in competitive contexts, students will not ask each other for help because to do so amounts to admitting a lack of ability (Ames & Ames, 1984). Ability grouping is an evaluative act that shows a competitive emphasis on ability and develops in the students a belief that ability is a linear factor. Without community, writers find it harder to develop an understanding of audience, and they get less evaluative feedback as they must depend on direct contact with the teacher.

The context of the evaluative transactions affects both the kinds of transactions and the way they are understood. For example, trust, or the lack of it, alters the whole game. Without trust, an apparently reflective queston like "What do you think?" can be heard by the student to mean: "Am I thinking the correct answer?" or, "What should I be thinking?" Even if the student hears the question as genuinely reflective, the overall classroom context determines what the student will reflect upon. The reflective lens we use is rather like a mirror. We can focus the mirror onlv on the part of us that causes us greatest shame, like the pimple on our nose or our untamed hair, and never the rest of ourselves. Or we might tilt the mirror and see a bigger, more distorted image. The focus of our reflection is determined in large part by the wav others have interacted with us. If, for example, in our writing the most prevalent response was to the neatness of our text, that is likely to be distorted into the largest part of what we see.

While it is seductively simple to base literacy evaluation on the easily measured subskill and workbook curricula of spelling tests, fill-in the blank worksheets, end of unit tests, and handwriting, those measures provide, at best, only limited and fragmented information (Brown, 1987; Wiggins, 1987). They do not tell us about the meaning being made by the students. In addition to providing limited information, such measures lead to the narrowing of both learning and instruction. A consistent emphasis on lower level, mechanical aspects of literacy learning (usually removed from meaningful context) becomes what students view as literacy learning. Students internalize our value system as it is reflected in our evaluations and use it to reflect upon and measure their own learning. Teachers, as learners, are no different. If test scores are the central evaluative feedback teachers receive about their teaching practice, that aspect will generally be distorted into the largest part of their practice that they see.

Evaluation and instruction cannot have separate goals. If our goal is to have students actively involved in the composition of meaning, the activities of the classroom need to provide students with opportunities to question, to rethink, to refine their thoughts, and to extend their understandings. Silent classrooms provide only limited opportunity for this to happen. Classrooms where students conference only with the teacher also provide limited opportunities. Language learning is social. Higher psychological processes are acquired through transaction with others (Vygotsky, 1978; Langer, 1989). When reading and writing are combined, teachers gain an appreciation of how children use them together, as the children develop a variety of powerful literacy routines for accomplishing various goals. When class is structured with time for students to discuss reading and writing, teachers have an opportunity to learn about student thinking and, ultimately, how to structure and support its development. Discussion, as large literate conversations, provides opportunities for students to learn diverse ways of thinking as they participate in classroom discourse that enables them to give voice to their responses, to support their interpretations, and to enrich their perspectives by becoming aware of the responses of others in their class (Langer, 1989). Teachers structure and provide demonstrations of what counts in literate discourse. But simply gathering a class together does not make discussion. Dialogue only occurs where it is clear that at least two positions are possible. Teachers who believe that there is a single right answer, or right interpretation will not get dialogue; they must live with monologue. A teacher's perspective notwithstanding, such a belief can also be developed by constant exposure to tests and worksheets that insist on a single correct option.

In other words, our approach to evaluation forms a context for the learning that takes place. As a counter-example to worksheets and testing, Brown and Cambourne (1989) have developed a retelling strategy, Read and Retell, that allows students to engage in intensive reading, writing, talking, and listening in a collaborative way while, at the same time, helping teachers learn about and evaluate the retelling texts the students produced and the reading and writing behavior of students as they produced the retellings. An integral portion of the procedure also involves learning from classmates. In the Read and Retell sessions students are first immersed in a genre or topic for several days. The actual plan (though there can be any number of variations) includes:

> *A time to predict* — on the basis of the title students do a written prediction—"What do you think this might be about? Why?" Students share predictions in small groups.
>
> *Everyone read* — the teacher first reads the text aloud, and then the students read the text silently as many times as is needed. The purpose is to understand, not to memorize. The teacher might stop after a bit and ask the students to consider their predictions. How do they hold up in light of their continued reading?
>
> *Retell* — without looking at the text the students spend about ten to fifteen minutes doing a written retelling. The retelling might be directed to a friend, an absent classmate, or a teacher to see how students respond to different audiences. It is important that the writer be able to read his retelling, but spelling and neatness are not of prime importance.

*Share and Compare* — in partners students compare their retellings for similarities and differences, as well as why they might have included or omitted any portions. Then students discuss with each other whether they thought the retelling confused, omitted, or changed anything that might have affected the meaning. Students also indicate what portion or feature they might borrow from the partner's retelling to include in their own and why they would borrow that particular portion.

Not only does the teacher learn about the students' development, but this kind of evaluation strategy serves a double function. One, students learn strategies (predicting, retelling) to help them understand texts. Two, they learn that there is more than one way to interpret a text and that it is acceptable to learn from others in a purposeful way that provides them with opportunity to read, to write, to listen, to speak, to think, to make decisions, and to explain and justify those decisions—using literacy to read, to write, and to reason. Notice, too, that in this approach, evaluation is inseparable from instruction, and reading is inseparable from writing.

## Developing Reflectiveness

The techniques we use for evaluation have, for us, an even more important effect. Central to becoming independently literate is the ability to self-evaluate, to be reflective. Without reflectiveness our students will develop a dependent and powerless literacy. Thus it is critical to realize that the ways in which we choose to evaluate have an impact on whether or not students choose to become reflective and the nature of their reflection. These effects are far-reaching and cyclical. When students do not self-evaluate, they cannot be independent in their learning. Student dependence restricts the kinds of evaluation and instruction a teacher can engage in.

Self-evaluation is central to being in control of one's own learning. When a teacher's responses indicate that she has all the knowledge and ownership of all the correct responses (and that there is only one for each situation) and that the student's job is simply to receive the knowledge, students do not learn to become reflective, to self-evaluate. In these classrooms students are more likely to develop an intellectual dependency and a different kind of literacy—one that controls rather than liberates. The focus of a teacher's evaluative responses is critical. When answers are simply right or wrong, a narrow form of evaluation occurs that leads to an even narrower form of learning. In these environments students see evaluation as something that is done to them. What self-evaluation they do engage in reflects how well they have measured up to the narrow criteria established by acquisition of the right answers and to other members of their class. In narrow evaluative situations students learn that it is the right answer, not the thinking involved in finding an answer, that is valued.

Yet children are capable of more than this. Jane Hansen (1987) found that even young students can be encouraged to evaluate themselves. Being able to self-evaluate puts students in control of their own progress, which is central to becoming independent learners. Such control is motivating and involving. Teachers who hold high expectations for their students believe that students can talk about what they know and give proof of their learning. The ability to do so confers security in a learner's knowledge. Students also learn when they need help and how to get the help. Learners know that making progress involves challenges. When students are encouraged to self-evaluate, not only do they learn to rely on themselves but they also learn that they are capable of finding many of their own answers.

Self-evaluation can be developed by encouraging learners to "have a go," to predict, and to check their predictions. Or students can be encouraged to consult peers or to use other resources in the room—what Hansen calls "independence through interdependence." In classrooms that produce this kind of learning, when students identify things they need help with, it is viewed as progress, not evidence of ineptness. Students are encouraged to view themselves historically by keeping records of their own learning. They are discouraged from viewing themselves against their peers as such comparisons do not foster cooperative learning and sharing, and such comparisons actively discourage asking for help.

Although we can help students view reading and writing in reflective ways, our evaluative efforts also help to determine what they reflect on. At a recent conference a speaker was recounting his own early academic history. He remembered always feeling like Patricia Riley Giff's character, Ronald Morgan, during his elementary years. He vividly recalled his classmate, Mary Margaret, the academic paragon of his class. In all their elementary years he could remember countless times when the exemplary Mary Margaret received lavish praise for her perfect work, but never once could he remember anyone asking her to talk about how she went about it. It was apparent to him that she was simply and innately "smart" and he was simply and innately "stupid." Such consequences can be reduced by supporting learners as they become self-evaluators by regularly asking students questions when they complete something or solve a problem like "How did you do that?" This helps students begin to think about how they go about things and to raise that to a conscious level. In the long run, this attention allows students to talk about the conditions and procedures for doing things that were successful or unsuccessful, which then provides a way to make informed decisions about the use of particular strategies in the future (Johnston & Winograd, 1985). And when classroom learning is social, this strategic knowledge can be shared. Indeed, in the classroom when we focus students' attention on the process more than the outcome and allow students to share their strategies with others, we decrease the likelihood of destructive comparisons while we increase the opportunity students have to learn from each other. Attending to the process is a fundamental descriptor of effective learners (Diener & Dweck, 1978). The evaluative transactions of the classroom simultaneously become both instructional transactions and ones that determine what kind of learners will become: external evaluation readily becomes internalized.

## Evaluation as a Means of Learning About One's Learning

Formal techniques for keeping track of students' development can have an impact on their reflectiveness just as much as informal ones. Reflection takes time, opportunity, trust, and patience. External accountability testing makes teachers feel that they have less time so that they cannot afford the luxury of patience or trust in their students. It drives us to require the right answer now as proof of the transmission of knowledge.

### *Learning Through Portfolios*

As a counter-example, part of students' evaluation may involve regularly putting together a portfolio of their work over a period of time. Portfolios have been used for evaluative purposes in a variety of classrooms from K–12 (Valencia, 1990), to college literature and writing classes (Elbow, 1986), to special education classrooms (Bartoli & Botel, 1988). Portfolios contain examples of the processes of the student's learning, as well as performance level exhibitions of student work. Process data might include drafts of a piece of writing, entries in a literature response journal, lists of questions the student had, rehearsals of all sorts. Product data might include the final draft of the written text, an interpretive reading from a student-selected passage from a piece of literature, a completed project related to the unit. Also included might be reflective comments on the pieces selected for the portfolio describing why they were selected—the growth they signify—and the process of how those pieces came to be.

The contents of the portfolio should relate to the goals for the particular instructional episode. For example, in a unit on folktales the teacher might have several goals; that:

- students become familiar with a certain number of folktales
- students have an opportunity to enjoy folktales in both written and oral form
- students can recognize the expressive language of folktales
- students can recognize and compare the characteristics of folktales
- students will be able to write their own folktale
- students will be involved in the active presentation of a folktale either through storytelling, reader's theatre, or music and dance.

A portfolio related to these goals might contain a certain number of the following:

- Selected samples of responses to folktales from the student's response journal
- The questions the student had about various folktales
- The drafts and performance copy of the student's original folktale

- A chart done either collaboratively or alone categorizing the characteristics of a selected number of folktales or a story map of a student-selected folktale
- A collage, or some other visual representation depicting the essence of a particular folktale
- An audio or visual tape containing the student's performance of a folktale.
- Process papers describing how particular pieces came about, the changes along the way, and the reasoning that went into the student's decision making.

Unlike more typical forms of assessment, the student would be expected to ask questions of both teacher and classmates, to receive coaching along the way, to view the doing as on-going learning, rather than simply as proof that the desired information has been transferred from teacher to student. Students would also have an opportunity to be involved in establishing the goals for the portfolio and the standards used for evaluating.

This involvement in their own evaluation produces an involvement in learning and a reflectiveness that is the hallmark of literate people. Involving students in the evaluation process in this way focuses the students' attention on their own growth, a way to view growth without shame. This historical account of student development also provides us with a way of seeing what students value and why. Jerome Bruner (1986) argues that much education comes from being able to distance oneself from what one knows—stepping back. When a learner distances herself, she is able to make connections, to think about what she knows and the method she has used, to manipulate that knowledge so that it can be used in other circumstances. Being mindful of what we do, and sharing this with others, fosters the growth of new ideas in the community.

## *Learning Through Interviewing*

Jane Hansen (1987) interviews students to help them develop as self-evaluators. She asked students four questions:

1. What's something new you've learned to do in writing?
2. What's something you would like to learn so you can be a better writer?
3. What's something new you've learned to do in reading?
4. What would you like to learn so you can become a better reader?

Hansen found that even young students had no difficulty evaluating what they could do and what they wanted to work on next. If we think that the prime purpose of evaluation is to provide optimal instruction for all students, we can see that the responses to the above interview questions can provide a wealth of information for making informed decisions for optimal instruction. Hansen discovered that many students, in response to those questions, also talked about what they do when they are reading and writing. In discussing their reading and writing processes, students reflected on how they have learned to learn that provides teachers with an opportunity to learn about the procedural knowledge the student brings to the literacy learning process. Other teachers have incorporated Hansen's questions into student evaluation folders. The folder is divided into three columns:

| Things I Can Do | Things I'm Working On | Things I Plan to Learn |
|---|---|---|

Students are thus capable of documenting and reflecting upon their own developing competence. This allows them to talk with certainty to parents and other interested parties about their learning and to become responsible for using their new knowledge. For example, if they now know how to use exclamation marks, they can be responsible for that aspect of editing their own writing. Nancie Atwell (1987) also interviews students in small groups as the center of her evaluation. Her secondary students discuss three things:

1. The main thing they hoped to learn
2. The main thing they did learn
3. One thing they intend to do because of what they learned.

She uses this documentation as part of her evaluation process and thus places a value on self-evaluation. Competence in self-evaluation grows over time. Teachers support students as they develop their ability to monitor, as well as the ability to step back and reflect on the learning that took place. Teacher demonstration and response help students to develop a trust in themselves as self-evaluators.

Often what students have learned goes unnoticed when reading and writing and thinking are seen as separate. For example, in interviewing a student recently we asked him what *he* thought was the most important thing that he had learned during a unit on plays. In this particular unit students had worked in small groups to read, rehearse, and perform short dramas. He responded that the most important thing that he had learned was ". . . that old people aren't really mean; they're just kinda set in their ways." He explained that in preparing for the part of an elderly man he found the character mean, but he just didn't understand why he was so mean. In his attempt to understand his character he listened in particular to how his grandparents talked to each other, and, still not satisfied, he spent time in a large mall sitting near various groups of elderly people and observing and listening to them. It seemed to him that people weren't being mean to each other, but they weren't really listening to what the other had had to say. In doing this he felt he came to understand his character better and also had learned something about older people. We often speak of literacy as having the power to inform and transform; interviewing provided an opportunity to learn about how and why this occurs.

James Squire (1987) argues that too many American students are simply unaware that they are learning anything and that by having students regularly ask themselves, "What am I learning?" students become conscious of their learning process. When we ask students to regularly think about what they've learned and what they are learning, we tell them, in effect, that they are capable knowledge makers. The attention to learning as an on-going process enables students to see that knowledge is not fixed but grows and changes over time. Teachers need not be the only ones to interview students; students can also interview each other as a regular complement to teacher interviews. The responses to student interviewers may provide additional perspectives on student learning (Farrell et al, 1988). As teachers we can also model the kind of self-evaluation that occurs in an interview by sharing with our students what we, too, are learning. By drawing on the information gained in interviews and through reflective writing, we can help students do with our help what they cannot yet do alone; as such, we join forces.

## *Learning from Students' Questions*

Student questions are not a major portion of the discourse of classrooms (Weber, 1986, Dillon, 1988). J. Dillon did extensive research on questioning and teaching. He observed 721 students in twenty seven classes and found only eleven student questions in all his observations. Students were the answer *givers* with teachers using this oral recitation as one way for students to show that the information had been transmitted from teacher or text to students. While correct response to known-answer questions provides some evidence of what students know, it provides limited insight into student knowing—the hows or whys they know what they know, what they wonder about and/or what they do not yet understand. It is that kind of inquiry upon which teachers help students develop their understandings.

Teachers, however, need not be mute members of literate communities. Langer (1988) looked at teacher questions in her research on how classroom discourse can support literate thinking. She found that teachers can ask questions that help students develop more expansive ways of thinking about reading and writing. For example, she found that the students demonstrated the richest literary thinking when they were treated as thinkers, assumed to have interesting and cogent thoughts about what they had read. Langer found that students used the most extended language and thought when they responded, not to known-answer questions, but to something she termed "student knowledge taps." These are questions that have no single right answer but which provide students with an opportunity to think and solve problems. Over time these knowledge taps became ways of thinking that students began to internalize. For example, the teacher might stop the students at a particular place in the reading of a text and ask the students what they were thinking and to make a prediction and discuss the basis for the prediction. Thus teachers, while learning about students' development, shaped the process of that development. Students also began to ask knowledge tapping questions of their own, such as, "But where will the food and fuel come from once they've used up all the supplies?" or, simply, "Why do I think this?" While there was no particular answer in the text, the students formulated and developed hypotheses of their own. In fact, students learned that having questions and thinking about the questions were important to their understandings. One seventh grade student described himself in the fall as a reader who focused on "facts." In an interview at the end of this year the student discussed his growth as a literacy learner.

> Now I think about all the things that could happen.
> It [a piece of literature] doesn't end at the end.
> . . . Just information doesn't help you. You'll always be reading more. Thinking about thinking when you read is something you take with you. You use it every time you read.

The knowledge gained from portfolios, exhibitions, interviews, and students' questions provides a rich source of evaluative information that would be trivialized by being reduced to numbers (Valencia, 1990). On the other hand, this is information that is easily shared with parents and administrators if periodic reports were to become humane dialogues rather than numeric printouts. In some schools where folders and portfolios are used, the faculty have become extremely frustrated with the limitations of standardized report cards. The teachers in one school district initiated narrative accounts describing student accomplishments over a marking period, although it was decidedly more work. Another school district initiated "talking report cards" where teachers guide parents through folders of student work. We think that there is no one correct response and that if teachers, administrators, students, and parents could discuss this, many useful alternatives to report cards would be discovered.

By contrast, normative assessment practices focus students' attention on their performance in comparison to others, and often on what they cannot do. Under such circumstances, students systematically ignore information about their own previous performance, deeming it irrelevant to their self-evaluation (Ames & Ames, 1984). Students in competitive contexts also tend to attribute the outcome of this comparative evaluation to an immutable ability. Normative testing can also have an impact on instruction, often focusing attention on fragmentary weaknesses—"needs more instruction on medial vowels." Normative assessments, because they are used to compare students across time and space, can create competitive contexts where students will not ask each other for help because so doing is seen as an indication of deficiency. When students view requests for help or questions as evidence of inadequacy that they must hide, real student learning, ultimately, will be vastly constrained.

## Teacher's Self Evaluation: Knowing Ourselves and Our Practice

Many of the issues we have raised about the consequences of classroom evaluation for children's learning can be directly translated to the context in which teachers operate. Teachers can be encouraged to be reflective or not, and the context of their reading of their students and of their own practice will change the interpretations that they make. For example, earlier in this chapter we quoted two teachers describing the literate development of students. We noted how teachers involved in that study who knew alot about literature generally reflected that knowledge in their descriptions of students' development. However, these reflections of teacher knowledge were not always evident. In one school district in which teachers were required to use basal readers along with all the tests and in which there was a heavy emphasis on testing, even teachers with substantial knowledge of children's literature did not reflect that in their descriptions of children's literacy development. Though they knew what to observe, they did not have the opportunity of observing the children selecting their books and their ways of responding to those books.

If instruction is to serve students better, teachers will need to become learners about their practice. We do know something about the conditions that foster and restrict such learning. A belief that there is one truly correct way to teach is just as destructive as a belief that there is only one meaning for a text. A competitive and individual context will produce restricted learning about teaching, just as it will restrict the learning about literacy.

The techniques used for keeping track of student development can make it easy or difficult for teachers to view their teaching practice reflectively. For example, the use of dialogue journals allows a teacher to review the history of her instructional transactions with students because the journal serves as an historical accounting of the written transactions between teacher and student. Journals provide one way for teachers to continue to make informed decisions about instruction based on what students are learning. On the other hand, the use of a basal reader substantially restricts the opportunities teachers have to observe children's literate activity and to reflect on their own instruction since the responsibility for instruction and planning is shifted from the teacher to the basal manual.

Certain more formal techniques for recording and evaluating oral reading behaviors, such as running records (Clay, 1985) and miscue analysis (Goodman, et. al., 1987), allow teachers to reflect on changes over time and also help them to develop further the ability to recognize patterns. This ability to recognize patterns is most strongly developed when teachers make a regular practice of engaging in detailed case studies with other teachers. In these discussions they hear other perspectives on students' patterns of behavior and, perhaps, learn alternate solutions to instructional problems. These "reflective conversations" among teachers (Carini, 1979) help teachers expand their ways of both looking and learning as they share the "reading" of children. Additionally, these reflective exchanges among teachers help them to keep reviewing the fine line between those aspects of the student's life and development that they can and cannot influence. It is this line that stands between teachers' burning out from accepting responsibility for things over which they have no control and their rationalizing a lack of control over things in which they could make a difference.

Checklists are one way to help teachers keep track of development. Human memory is frail, and teachers have a lot to remember. Checklists can help ensure that students do not fall through the cracks. For example, they can be used to keep track of which students have been seen in conference recently and which have not. Checklists of concepts and strategies can also be helpful for keeping track of student development. Checklists can also help monitor the amount and kind of students' participation in group discusson. The possible list of items that might appear on a checklist can become burdensome. Decisions about what might appear on a checklist are important because they represent a valuing of specific behaviors or knowledge. When teachers get together to compose a developmental checklist, important things happen. Teachers debate what should and should not be on the list in order to obtain the balance between functional length and critical items. In that dialogue, teachers have an opportunity to learn from each other, to have an investment in the recording procedure, and internalize most of it at the same time. But checklists, like report cards, should also become the subject of teachers' revision and reflection. Does the checklist reflect what is valued in our classrooms? Is it reflective of the opportunities to learn in our classrooms?

When we take time to be analytical about the learning in our classroom, we look again, and this re-vision allows us to evaluate the student learning and our own pedagogy (Lincoln & Guba, 1983; Harste, Short & Burke, 1988). A first grade teacher talked about establishing a formal routine for looking again at what her students could do. Every other week her students select a piece of their writing for her to copy. She analyzes each piece, looking for what the student can do. She makes note of this on the paper. She also compares the current work with previous work. It does not take an enormous amount of time—about an hour—but what she learns from this periodic analysis gives her not only a greater and more tangible record of student development, as a guide to providing more effective instruction when and where it is needed, but also serves her as an opportunity to reflect on the kinds of support she has provided for these young authors and its effect.

Those who teach teachers in undergraduate and graduate level courses need also to be reflective about that preparation. Too often teachers are prepared to be only spectators, rather than participants, at literacy events. We think that it would be useful if those preparing to be teachers had greater participatory roles in literate communities: or if, for example, they had experience keeping their own response journals in the courses they took; if they had opportunities to be readers and writers and not just to learn about writing and reading from an outside perspective; if they might put together their own portfolios, representing both singular and collaborative work, and have greater involvement in reflective conversations (Carini, 1979) and self-evaluation. In a NAEP report (Applebee, et al, 1989) findings indicated recent emphasis on processes involved in reading and writing have not been positively related to achievement. The NAEP report suggests that perhaps process reforms have been implemented in classrooms in a superficial way or in ways divorced from actual reading and writing. Similarly Sarason (1971) reminds us that in terms of educational reform

> If the more things change, the more they remain the same, it is because our ways of looking and thinking have not changed.

If we desire to use evaluation to inform more optimal instruction for all children, teachers must change their ways of looking. This is most likely to happen if we help teachers become active learners of literacy, as well as their own practice. Certainly teachers can and are doing this on their own, but there is greater potential for long-lasting change if teachers as learners had the opportunity to participate in their own literacy development before we expect them to observe the development of others.

## Conclusion

Evaluation, we believe, is not something final or fixed. It is, rather, something on-going that both students and teachers do to improve learning and teaching. The forms of evaluation that teachers use provide ways of thinking that our students internalize. Evaluation practices that help students understand and recognize their developing knowledge and competence serve a generative function for both learning and instruction. Certainly, not all problems in literacy learning are eliminated by the kinds of evaluation discussed in this chapter. But it is our belief that when we develop ways to know the individuals that we teach, we become more involved with students' learning and students, in turn, become more involved with their own learning. Evaluation is central to teaching and learning. Teachers, in their day to day transactions with their students, exert a powerful influence on how children will come to evaluate and monitor their own learning. To ignore this serves no one whose goal is optimal literacy learning.

## References

Ames, C. & Ames, R. (1984). Goal structures and motivation. *The Elementary School Journal*, *85*, 39–52.

Applebee, A.N. & Langer, J.A. (1983). Instructional scaffolding as natural language activities. *Language Arts*, *60* (2) 168–175.

Applebee, A.N., Langer, J.A. & Mullis, I.V.S. (1989). *Crossroads in American education*. Princeton, NJ: Educational Testing Service.

Atwell, N. (1987). *In the middle: Writing, reading, and learning with adolescents*. Portsmouth, NH: Boynton/Cook, Heinemann.

Bartoli, J.S. & Botel, M. (1988). *Learning disability: An ecological approach*. New York: Teachers College Press.

Bissex, G. (1980). *Gnys at wrk: A child learns to write and read.* Cambridge, MA: Harvard University Press.

Brown, H. & Cambourne, B. (1989). *Read and retell.* Portsmouth, NH: Heinemann.

Brown, R. (1987). Literacy and accountability. *Journal of State Government, 60* (2). 68–72.

Bruner, J. (1986). *Actual minds, possible worlds.* Cambridge, MA: Harvard University Press.

Carini, P.F. (1979). *The school lives of seven children: A five-year study.* Grand Forks, ND: North Dakota Study Group.

Clay, M. (1985). *The early detection of reading difficulties (3rd ed.).* Portsmouth, NH: Heinemann.

Diener, C.I. & Dweck, C. (1978). An analysis of learned helplessness: II. The processing of success. *Journal of Personality and Social Psychology, 39,* 940–952.

Dillon, J.T. (1988). *Questioning and teaching: A manual of practice.* New York: Teachers College Press.

Elbow, P. (1986). *Embracing contraries.* New York: Oxford University Press.

Farrell, E., Peguero, G., Lindsey, R. & White, R. (1988). Giving voice to high school students: Pressure and boredom, Ya know what I'm sayin? *American Education Research Journal, 25,* (4), 489–502.

Goodman, Y.M., Watson, D. & Burke, C. (1987). *Reading miscue inventory: Alternative procedures.* New York: Richard C. Owen.

Hansen, Jane (1987). *When writers read.* Portsmouth, NH: Heinemann.

Harste, J.C., Short, K.G., & Burke, C. (1988). *Creating classrooms for authors.* Portsmouth, NH: Heinemann.

Johnston, P. (1987). Assessing the process and the process of assessment in the language arts. In J. Squire (Ed.) *The dynamics of language learning.* Urbana, IL: National Council for Research in English.

Johnston, P.H. & Winograd, P.N. (1985). Passive failure in reading. *Journal of Reading Behavior, 17,* 279–301.

Langer, J.A. (1989). *The process of understanding literature.* (Report 2.1). Albany, NY: Center for the Learning and Teaching of Literature.

Langer, J.A. (1988). Sociocognitive perspective on literacy and language. In J.A. Langer (Ed.), *Issues of society and schooling.* Norwood, NJ: Ablex. 1–26.

Lincoln, Y.S. & Guba, E.G. (1983). *Naturalistic inquiry.* Beverly Hills: Sage Publications.

Rosenblatt, L.M. (1984). *Literature as exploration* (3rd Ed.). New York: Modern Language Association.

Sarason, S. (1971). *The culture of school and the problem of change.* Boston, MA: Allyn & Bacon.

# IV-10

## Using Assessment as a Tool for Learning

JUDITH A. ARTER*
NORTHWEST REGIONAL EDUCATIONAL LABORATORY

Mention the word assessment to most students and teachers and they will immediately respond with the concept of "tests"—the kind used to monitor and track student performance—TESTS—the type given to produce grades, select students for special programs, and report to the board of education. Wouldn't it be exciting, however, if we associated the word assessment with a process that (a) provides an opportunity for teachers and students to thoughtfully examine what has been learned, how we know it has been learned, and what needs to happen next to reach a new level of learning and (b) is an episode of learning in and of itself?

Assessments can be a significant tool in the instructional process, one that can demystify the learning process so that teachers and students become equal partners and shareholders. Assessments can also become a tool for learning in and of themselves; students can learn something from doing them. In this arena, educators do not teach and then assess; nor do they think of assessment as something that is done *to* students. Instead, they consider the assessment activity itself an instructional episode. Let's look at some examples of such assessments.

### Performance-Based Writing Assessment

Performance assessment is "assessment based on observation and judgment" (Stiggins 1994): we observe a behavior or examine a work product and judge its quality. This kind of assessment is frequently done in the classroom by teachers and is common in daily life as well (e.g., job performance reviews, driver's tests, the Olympics). If done properly, performance assessment has the powerful and enabling quality of being a tool for learning as well as a tool for generating numbers on students.

The *criteria* we develop to judge the quality of student work are critical to making performance assessment a powerful tool for learning. When we go through the process of examining student work samples to document varying performance levels, we create a set of criteria that can be used to judge the quality of the performance. The criteria, however, can be more than just a mechanism to score or grade performances; they can also clearly communicate what is valued in the work—what it takes to be successful. Criteria are teachable, and they provide students access to the "secrets" of success (in fact, students can be full partners in developing criteria). Thoughtfully written and teachable criteria support our moral obligation to make progress visible to students. When students and teachers begin to share the same vision of success, they quickly develop a vocabulary for discussing this vision. All are working toward the same clear and identifiable target. This improves student achievement.

---

*With special thanks to Ruth Culham for editorial assistance.

To get some understanding of how this works, let's analyze two samples of student work. Read the two student papers that follow ("Redwoods" and "Fox") and take a moment to jot down your reactions to them. What did you notice? Are there any obvious strengths? Are the strengths the same in each paper? What features did you respond to? Which paper did you like best? Why?

### *The Redwoods*

Last year, we went on a vacation and we had a wonderful time. The weather was sunny and warm and there was lots to do, so we were never bored.

My parents visited friends and took pictures for their friends back home. My brother and I swam and also hiked in the woods. When we got tired of that, we just ate and had a wonderful time.

It was exciting and fun to be together as a family and to do things together. I love my family and this is a time that I will remember for a long time. I hope we will go back again next year for more fun and an even better time than we had this year.

### *Fox*

I don't get along with people to good, and sometimes I am alone for a long time. When I am alone, I like to walk to Forests and places where only me and the animals are. My best friend is God, but when I don't believe he's around sometimes, My dog stands in. We do everything together. Hunt, fish, walk, eat and sleep together. My dog's names is Fox, 'cause he looks like an Arctic Fox. Fox and I used to live in this house with a pond behind. that pond was our property. the only thing allowed on it (that we allowed) was ducks and fish. If another person or dog would even look like going near that place, Fox and I would run them off in a frenzy. there was a lot of rocks around, so I would build forts and traps for anybody even daring to come near. The pond had a bridge that was shaded by willows, so on a hot day me and Fox would sit on that bridge and soak our feet, well, I would soak my feet, Fox just kinda jumped in.

At night, the pond was alive with frogs, so I would invite this kid over, (he was a guy like me) and catch frogs. After we had a couple each, we would pick the best looking one out of our group and race them. The winner gets the other guys Frog.

In the winter, the pond would freeze over, and I got my ice skates out. the pond was now an ice skating rink. Fox would chase me as I went round and round the pond.

After about a year, I was riding my bike patroling the area around the pond. With Fox at my side, I raced downhill toward the pond. I tried to stop, but my back tire went into a skid. I went face first into murky, shadowy waters. When I went down, a minute later I felt something pull on my shirt. I grabbed it, not know what to think. When I hit the surface, I saw that it was Fox, pulling on my shirt as if he was trying to save me. He was to little to save me if I was really drowing, but it was the thought that counts. I owe him one.

Another year passed. One day my mom got home from the store, and she bought me a rubber raft. It was just a cheap one, but it was mine. I blew it up with a tire pump. It was just the right size for me and fox. Out of respect for Fox, I named it the USS Fox and chirstened it right in the pond.

On sunny days, I would take the raft out and lay in the sun with Fox on my legs.

One day, when I was asleep in the raft, the wind blew pretty hard and blew my raft right into a bunch of sticks and rocks. The USS Fox was given a sad salute, and then was no more.

Another year passed, and this would be our last year by the pond. I admired and respected that pond more than I ever did that year. But, at long last, all good things must come to an end, we moved to another town. Fox and I still visit the pond, but it'll never be like them 3 years when she was mine.

Most teachers and students reading these papers identify the same sorts of things as features they like and dislike. In "Redwoods," an 11th grade paper, these comments are typical: vague, not focused, unclear picture, very general, technically correct (capitals, spelling, punctuation, etc.), safe paper, and uninteresting. Reactions to "Fox," an 8th grade piece, are quite different. Teachers and students often make comments like these: focused and clear, I felt like I was there, there's a natural rhythm and flow to the language, creates a picture, nice use of words, good examples, good beginning and end, personal reflection, original, descriptive, purposeful, pulls from life, exciting, nice pacing, humorous, engaging, vivid, interesting, emotional, and revealing. How did your list of comments compare to these?

## Six-Trait Analytical Model

Comparing student writing with such different attributes helps us to sort through the complexity of features that make writing successful. Such exercises can help teachers and students explore what is valued in writing. In fact, the same features recur so often on brainstormed lists from teachers across the country that they form the basis of a set of performance criteria for writing called the Six-Trait Analytical Model. "Six traits" because teachers' comments seem to group themselves into six areas: ideas, organization, voice, word choice, sentence fluency. and conventions. In other words, we tend to value more highly writing that has clear and insightful ideas, is organized in such a way to support the ideas presented, has an honest and personal voice, uses words well, is fluid and easy to read, and is conventionally correct. Look at the way teacher comments fit into the six traits (see Figure 1). Do yours fit in the same way?

The six-trait model for analyzing writing is basically an attempt to systematically write down what we already value in writing so that we can be consistent in our vision of success across students, across assignments, and across time—and, more importantly, so that students can share in the vision of success and not have it constantly shifting in unknown ways between teachers or assignments.

FIGURE 1

WHAT TEACHERS OF WRITING VALUE

**Ideas**
- Reflects real life
- Clarity
- Narrowed, focused, manageable topic
- Quality details you care about
- Makes you think

**Voice**
- Writing that takes a risk
- Personal flavor, tone, style
- Commitment to the topic
- Sense of involvement
- Emotion—humor
- Sounds like a person wrote it

**Sentence Fluency**
- Rhythm
- Cadence
- Variety in sentence beginnings
- Sentence beginnings that link ideas
- Fragments add style
- Dialogue rings true
- Fun to read aloud

**Organization**
- Opening that hooks you
- Sequencing that makes sense, holds your attention
- Coherence
- A thoughtful wrap-up
- No "dream" endings
- No ho-hum summaries

**Word Choice**
- Words and phrases you love and remember
- Word "pictures" you can't forget or ignore
- Phrasing with a personal touch
- The familiar word in new surroundings

**Conventions**
- Generally correct
- An eye on the details
- Looks proofread—edited
- Editing enhances readability
- No "tear your hair out" errors
- Reading is a pleasure

*Vicki Spandel and Ruth Culham*
*Northwest Regional Educational Laboratory, Portland, Oregon*

### Value of Criteria to Teachers

Have you ever been nervous grading student essays? Those who have rated or scored student essays for a formal writing assessment, using criteria such as the six traits, unanimously agree that the very process of assessing teaches a great deal about what makes writing successful (as we discovered above). Raters have to internalize a set of criteria so thoroughly that they can be consistent both with themselves over time and with other raters. During the process of scoring hundreds of papers, raters see many examples of what good and poor writing looks like and they learn to systematically analyze it to determine why it is good or poor. Anecdotal reports from raters indicate that, for many, this assessment experience marks the first time they truly feel confident that they can identify the features that make writing work, in large part because they have a systematic vocabulary for talking about these features. These teachers go back to the classroom much more confident in commenting on student work and helping students understand what it means to write well. Equally important, raters say they become better writers as a result of scoring students' writing.

Great! As teachers become more comfortable with their own writing and more consistent in their responses to the features of good writing, they gain the confidence to share their insights with students. If assessing brings about a shared vocabulary and a deeper sense of what it means to write well, then shouldn't students also benefit from the same process? The answer is YES!

### Value of Criteria to Students

Students can also learn to analyze sample papers, develop a systematic vocabulary, and put criteria into their own words so they can develop their own sense of what it means to write well. To illustrate the need for this process, consider peer-response groups in writing. For a long time now, teachers have been concerned that during peer-response time students tend to give each other superficial feedback such as "I liked it" or "It was good," Maybe one reason for this superficiality is that students don't know what to look for and have no common vocabulary for framing their comments. I suggest that if we systematically help students understand the relevant dimensions of a performance, help them recognize what works, and help them develop vocabulary to discuss what they see, their comments will become much more sophisticated.

BEST COPY AVAILABLE

> I want (students) to see evaluation in its best sense—a source to inform teaching and learning. To that end, we develop a vocabulary for commenting on the admirable and problematic aspects of writing. . . . The more we examine samples, the richer and more helpful this language of evaluation becomes (Erikson 1992).

The goal of assessment should not be just to earn grades. The goal of assessment should be to help students know how to do it better next time. Clear performance criteria have this effect.

> "Winning points" may be the final goal of classroom work as it is of the sports endeavor, but the grade, like the final score of the game, never taught anyone how to win again, or why they lost. For the truly successful contenders, playing the game is always about learning the game . . . however often it seems to be about scoring more wins than losses" (Lucas 1992).

> Scales, criteria, and specific questions that students apply to their own or others' writing also have a powerful effect on enhancing quality. Through using the criteria systematically, students appear to internalize them and bring them to bear in generating new material even when they do not have the criteria in front of them (Hillocks 1986).

This is instructional assessment. Students are learning something in the process of assessing and being assessed. They are learning about writing and then applying this knowledge to their own and others' works. They learn what effective writing looks like and what to do to make their own writing stronger through the revision process.

## Beyond Writing Assessment

Good performance criteria are more than just a tool that teachers (or districts or states) can use to monitor students' ability to write or solve problems or be critical thinkers. They also help teachers and students conceptualize and define standards, especially in hard-to-define areas such as critical thinking, problem solving, collaborative learning, and communication. They provide a vocabulary for discussing features of work or behavior that are subtle and ill defined. Good performance criteria help teachers answer such questions as these: "What is expected?" "What are our standards?" "What does good performance look like?" "What do I want to accomplish?" "What kind of feedback do I give to improve student work next time?" They help students to answer similar questions: "What is expected?" "What does good performance look like?" "How will I know when I'm successful?" "Why did I win?" "How can I win again?"

### Establishing and Having Clear Criteria

There are two essential features here: (1) having a clear picture in our own heads of the criteria for success, and (2) letting students in on it.

**1. Having a clear picture in our own heads of the criteria for success—or—performance criteria clarify the goals we have for students.** They are, in effect, the final and most complete definitions of district and teacher goals for students. Notice that the performance *criteria*, not the performance *tasks*, have this function. Performance criteria define what we value in student work, define our targets and goals, and are useful in determining when we get there.

In many discussions of integrating assessment and instruction, criteria are left out. There is lots of talk about the *tasks* that should be given to students and how one watches performance on these tasks to monitor where students are. But the question is, what exactly do you look for in the performance in order to determine progress, know when students are ready to move on to the next topic, or determine when mastery has occurred? Criteria tell us what to look for. Overt and public criteria, arrived at through discussion and collaborative learning and accessible to both teachers and students, bring judgmental assessment from the realm of individualistic impressions to the realm of systematic and generalizable documentation of learning. They also result in assessment that promotes learning. (Not all performance criteria are created equal, however; all do not have this effect. Certain design features make them more or less useful as instructional tools. These features are outlined in a related paper by Arter in Section VI.)

**2. Letting students in on it—or—students must be involved in assessing.** Students learn to the extent they are the ones doing the judging:

> By encouraging students to engage in self and peer evaluation teachers empower students to take control of their own learning. When students help determine the criteria for assessment they can make reasonable decisions about the quality of their own work. By engaging students in self-assessments, students learn they are ultimately responsible for their own learning. It is in this area that portfolios are so powerful. (Tierney 1991, p.59)

As a matter of fact, one perspective is that engaging students in conscious development and systematic application of criteria is a major attribute of critical thinking. Therefore, the process of engaging students in assessment actually helps them to attain some of the thinking goals we have for them.

> To be successful, students must learn how to evaluate rather than how to be evaluated. . . . Evaluation is routine, something to be done

> at every step of the way as they examine decisions, observations, data, ideas and outcomes. . . . [When] evaluation is a key process, and one over which students have considerable control, they are truly critical thinkers (Penick 1989).

### Relationship Between Classroom Performance Assessment and Large-Scale Assessment

So far, it may seem that we are discussing only classroom performance assessment. The whole enterprise is much more powerful, however, if large-scale and classroom assessment are considered as a unit. Consider the following points:

1. What if performance criteria for large-scale assessment (which have to be developed anyway) were to be developed by teachers with classroom uses in mind? To develop performance criteria, teachers must engage in the discussions and do the research necessary to develop a common vision of what it means for students to be successful. Therefore, the effort that goes into large-scale development is also staff development and the development of instructional materials.
2. What if the performance criteria were uniform across teachers, so that we could be confident that, regardless of who observes a student performance, the judgment would be the same? Then assessment credibility improves—and students get the same message across teachers and grades. Large-scale assessment development can take performance assessment out of the realm of individual interpretation and can promote essential dialogue about what, in fact, constitutes our vision of success for students.
3. What if teachers scored large-scale performance assessments Then the effort that goes into scoring is also staff development.

### Test Specifications

Even though the above example is from the area of performance assessment, I believe other kinds of classroom tests and assessments can also be made more instructional. Because of the instructional demands on teachers, many times we unconsciously use the tests embedded in instructional materials, assuming that they will measure what is important. It is useful to bring the content of traditional tests to consciousness and to plan them rationally.

One way to do this is through the use of a "table of specifications" (Stiggins 1994), which is a blueprint for a test that clearly outlines the instructional targets for a unit and specifies how each target will be measured. For example, the sample table of specifications in Figure 2 is for a unit on pollution. The teacher has four types of information for students to learn in this unit (listed in the left column). The teacher also wants students to manipulate this information in various ways, from factual recall to critical analysis (listed across the top row). The cells in the table specify the methods for assessing whether students have attained the goals of the unit. For example, one type of information the teacher wants to cover is solutions to pollution. He wants the students to (1) learn some solutions and recall them (three multiple-choice questions), (2) make some inferences about solutions (why such a solution might have been chosen, three multiple-choice questions), and (3) evaluate the potential success of a proposed solution to a pollution problem (one essay critiquing a solution strategy). The numbers in parentheses indicate the approximate time that it will take students to answer the questions described. In the column totals, it is clear that recall, inference, and evaluation are of equal weight; in the row totals, solutions to problems are more important than the other topics. Now, how could such a table be used as an instructional tool? Here are some ideas:

FIGURE 2
TABLE OF SPECIFICATIONS

| | **Recall** | Inference | **Evaluation** | **Totals** |
|---|---|---|---|---|
| **Types of Pollution** | 3 m.c. | | | 3 m.c.<br>(3) |
| **Sources of Pollution** | 1 matching (10 items)<br>(3) | 3 m.c. (what probably caused ...)<br>(3) | | 10 matching<br>3 m.c.<br>(6) |
| **Effects of Pollution** | 5 m.c.<br>(5) | 1 essay (predict the effect of ...)<br>(10) | | (5 m.c.<br>1 essay<br>(15) |
| **Solutions** | 3 m.c. (recall some solutions)<br>(3) | 3 m.c. (why might ...)<br>(3) | 1 essay (evaluate the feasibility and effect of a solution)<br>(15) | 6 m.c.<br>1 essay<br>(21) |
| **Column Totals** | 11 m.c.<br>10 matching<br>(14) | 6 m.c.<br>1 essay<br>(16) | 1 essay<br>(15) | (45) |

m.c. = multiple choice questions

1. Develop the table of specifications in advance and share it with students so they know what is expected.
2. Modify the table during instruction (with input by students) if the relative importance of topics or other things change.
3. Have students write the questions (plus scoring guides) that match the specifications. To do this, they need to know what *inference* and *evaluation* are and how to elicit this type of response.
4. Have students predict how they will do on each topic, type of thinking, or item types, and then compare their predictions to what actually happens.
5. Have students critique the questions on the test. Did they really match the specifications?

In these ways, students actually learn something from the test; the test enhances learning.

## Conclusion

Common features of these examples characterize assessments that are also instructionally useful. First, each of the examples helps us to clarify the targets we have for students, both in our own minds and in students' minds. They help to answer questions such as these: What do we want to accomplish? Where are we going? What does good look like? How do we know when students are progressing satisfactorily?

Second, each of the examples encourages both teachers and students to reflect on their work. Assessment is not something done *to* students; rather, students are involved in defining targets and assessing their own progress toward them.

Finally, assessment has direct links to instruction. In fact, instruction is organized around the concepts in the assessment; likewise, assessment is organized around the concepts in the instruction.

If we truly want to ensure that all students attain the complex skills and thinking processes that will enable them to be successful and productive citizens, we must ensure that we know what success in these complex targets looks like and what constitutes satisfactory progress toward the ultimate target. Attention to the careful design of assessment forces us to confront these definitional issues. Viewed in this manner, good assessment becomes not just an exercise in monitoring, set apart from instruction, but the essential ingredient that forces us to be clear about what we want to accomplish with our students. It also becomes a tool for helping students attain those goals. Good assessment is an exercise in clear thinking.

Our assessments communicate what we value. If we value student ownership of learning, critical thinking, sharing classroom power with students, and skills that go beyond the regurgitation of other people's knowledge, we must begin to design and use assessments as tools for learning.

### REFERENCES

Erikson, M. (1992). "Developing Student Confidence to Evaluate Writing." *Quarterly of the National Writing Project & The Center for the Study of Writing and Literacy* 14, 1: 7-9.

Hillocks, G. (1986). *Research on Written Composition.* Urbana, Ill.: National Council of Teachers of English.

Lucas, C. (1992). "Introduction: Writing Portfolios—Changes and Challenges." In *Portfolios in the Writing Classroom,* edited by K. Yancy, pp. 1-11. Reston, Va.: National Council of Teachers of Mathematics.

Penick, J. (November 1989). "Issues-Based Education—A Need and Rationale." *Education and Urban Society* 22, 1: 3-8.

Spandel, V. and R. Culham. (1994). *The Student Friendly Guides to Working with Traits.* Portland, Oreg.: Northwest Regional Educational Laboratory.

Stiggins, R. J. (1994). *Student-Centered Classroom Assessment.* Columbus, Ohio: Macmillan.

Tierney, R. J., M. A. Carter, and L. E. Desai. (1991). *Portfolio Assessment in the Reading-Writing Classroom.* Norwood, Mass.: Christopher Gordon, Publishers, Inc.

# IV-11

# Collaborative Assessment: Making High Standards a Reality for All Students

DORIS SPERLING
ANN ARBOR, MICHIGAN

Teachers too often express criteria for successful work in vague terms with little meaning for students; this problem pervades classroom teaching and hinders efforts to help students meet higher standards. To improve educational performance, educators must aim to raise standards and overcome our failure to communicate to students exactly what these standards are and how we can help all students reach them.

This paper illustrates how a new approach to performance assessment, which I call "collaborative assessment," can make a major impact on the capacity for classroom teachers to help all of their students meet higher educational standards.

## What Is Collaborative Assessment?

The core of collaborative assessment includes these four elements:

- Teachers use their expertise to develop and articulate clear and completely understandable criteria to students.
- Teachers create the opportunity for students to participate in the process of developing the criteria.
- Because the criteria are clearly spelled out, students can take responsibility to evaluate their own work, compare their self-assessment with the teacher's assessment, set goals for future work, and initiate corrective action to improve their own work.
- Because teachers can observe the evolving instructional needs of the students, and because students become critically aware of their own strengths and weaknesses, teachers and students can collaborate regularly to refine, improve, and add to the assessment criteria.

A main benefit of collaborative assessment, shared by other forms of performance assessment, is that teachers must clearly define and communicate to students the "targets" they need to hit to succeed. However, collaborative assessment offers additional and compelling educational benefits.

First, when students are brought into the process of developing the full range of performance criteria—from poor to excellent—they gain a richer appreciation of what the standards for success are.

Second, collaborative assessment creates a classroom environment in which students have the requisite tools for judging the worth of their efforts and the avenues to self-improvement; they are empowered to take control of their own learning.

Third, collaborative assessment gives teachers specific clues about how they can help students learn. In addition, with students as ongoing assessors, they too take an active role in determining their instructional needs.

Finally, collaborative assessment allows others, most importantly parents and peers, to share in the assessment/instruction cycle.

As the Classroom Assessment Specialist for the Ann Arbor Public Schools, I have helped many teachers put collaborative assessment to work and witnessed firsthand the profound impact it has had on improving student achievement at all levels. Although we have not yet performed any controlled studies, teachers using collaborative assessment have documented that almost all of their students have shown extraordinary improvement.

## Collaborative Assessment: A Case Study

Collaborative assessment can be used in any subject and at any grade level. The case study I have chosen involves the writing program of Gail Hughes, a 4th grade teacher at Carpenter School in Ann Arbor.

Mrs. Hughes approached me expressing dissatisfaction with her students' progress. Her writing program involved the students' participation in a "Special Person" writing activity. Every week, a different member of the class answered questions about his or her personal life, interests, and ideas: this person was designated the "Special Person."

Students then wrote a summary of the Special Person's significant ideas, and these compositions were saved in each student's own portfolio. This format offered Mrs. Hughes a unique opportunity to examine each student's writing over a period of time. After several months, she felt the quality of writing had shown little or no improvement and, after examining the students' work, I agreed.

Despite Mrs. Hughes' efforts to teach her students how to make their writing more expressive, thoughtful, and effective, most students continued to simply list the Special Person's oral responses, with little sense of such things as author's voice or descriptive expressions. This pattern did not significantly change from week to week. Here is one girl's paper, which is representative of the class compositions:

> Andrea is a neat person. She likes rock bands. Andrea thinks reading is cool. Her favorite books are "Little House on the Prairie and "J. T." She has a cute dog that barks "Rudolph the Reindeer." Andrea and her family enjoy going out to eat. She likes shopping at Briarwood Mall. She thinks "Home Alone" and "Hook" are neat movies. Andrea is a very neat person and I'm glad I know her.

### The Teacher Develops and Articulates the Criteria

In reviewing the graded papers, I could not figure out what criteria Mrs. Hughes was using. Similar papers seemed to have been given different grades. When I asked Mrs. Hughes for her criteria, she realized that she had none in writing and that she had difficulty even verbalizing the criteria.

Mrs. Hughes quickly recognized the need to define her criteria. The process was indeed a challenge. She tried to use clear and specific language so that students would know exactly what was required to achieve the desired results. Her struggle to avoid such vague phrases as "thoughtful, expressive, and effective" was an ongoing effort. Even when Mrs. Hughes finally arrived at a satisfactory set of criteria, we anticipated there should be changes in the future.

### Students Participate in the Development of the Criteria

With well-defined criteria in mind, Mrs. Hughes now drew the students into the collaborative assessment process. She wanted her students to go through the same analytical process of determining criteria for high-quality writing as she had.

She developed four writing samples, each having the attributes of one of four levels (✓-, ✓, ✓+, and A) and gave them to her students to grade using their own best judgment. Here are the four samples, from the poorest to the best.

> 1. Sally likes to go to college she would like to go to Hellman College or U. of M. Sally wants to have a happy life she wants a boy and a girl she wants to travel to London and France she want to remember everybody in our class she wants to be rich when she is older she wants a black sports car.
>
> 2. Her goals are to go to college. She would like to go to Hellman or U. of M. When she grows up she wants to be an actress or a model. She would travel to England, London and France. She want to have a happy life with her family. She would want to have a boy and a girl. She want to be kind and gentle, not a bossy person. She wants a sports car. She want to live in Ann Arbor in a big house. She wants a lot of money. She wants a robot to clean her house. Her problem will be to stop endangered species.
>
> 3. Sally hopes her future will be wonderful. She would like a husband with a good job and two kids, a boy and a girl. U. of M. and Kalamazoo College are the colleges Sally might like to attend. There are so many problems in the world. Out of all of them Sally would like to help endangered species. If Sally could travel anywhere she would travel to England. The city Sally would like

> to visit in England is London. I wonder if Sally would like to visit the queen of England?
>
> 4. Sally hopes her future will be great. Probably her biggest goal is to attend college. The colleges she is thinking about are Kalamazoo College. I think she likes the city of Kalamazoo. Other choices are U. of M. or Hellman College. One of her friends went to Hellman and told her it was an excellent choice.
>
> She would love to travel to England and France. I think she might like to speak French if she goes to France. Sally wants to have a spiffy black sports car. Oh, I forgot. She would like it to be a convertible.
>
> She would like to be the kind of person who never bosses people around. Having a robot to clean her house would top off her life and make it perfect.

The students (who of course did not get these examples in the above order) worked in groups to assign grades to each selection. They were then asked to explain why they thought one paper was better than another. Using this method, Mrs. Hughes guided the students to think about and contribute to the writing criteria.

After incorporating the student input, she noted their constant use of the word "interesting" and recognized it as an example of vague criteria. Consequently, Mrs. Hughes developed the concept of "+ sentences" to define criteria for the word "interesting." She included such attributes as humor, author's voice, personal opinion, colorful adjectives, similes, and own observations.

Many models were provided to demonstrate what "interesting writing" looks like, including examples from literary authors. Mrs. Hughes also challenged students to select "+ sentences" from their own work to read to the class. This modeling technique was an important tool for helping students recognize good writing.

### The Assessment Process

After extensive work, teacher and students produced a written list of criteria. Using this list, I designed an assessment form to be used by each student and the teacher to evaluate how well the student was performing. The form, shown in Figure 1, consists of a "performance continuum" that describes the characteristics of papers ranging from poor to excellent.

Students must have a clear picture of what is unacceptable as well as what is acceptable. When students see that their writing resembles the unsatisfactory model, they become highly motivated to improve. With the criteria in the continuum providing a clear road map, students know specifically what to do to be successful.

I will never forget the 3rd grade boy who, when using a performance continuum for the first time and seeing how close his paper was to the poor criteria, asked permission to redo the paper. The paper he brought in the next day (and his assessment of it) were far closer to the excellent end of the continuum— and included his notation that from then on he was going to get all *A*'s. Having to judge his own work for the first time opened his eyes to the reality that he was capable of doing much better.

In Mrs. Hughes' class, students' assessment forms were on their desks as a resource every time the students wrote. Each week students also completed the self-assessment portion of the form, judging their own writing performance based upon the assessment criteria.

The weekly papers and the students' self-assessments were turned in to the teacher, who wrote her evaluation on the same form and returned it before the next week's Special Person writing activity. Students studied any difference between the teacher's assessment and their own. When necessary, individual conferences were held. With the assessment of their past week's writing clearly in mind, students were asked to set a personal goal to improve the new week's activity.

### Collaborative Assessment: Refinement, Additions, and Reflections

As predicted, the assessment form needed to be changed to reflect the instructional needs of the students. For example, Mrs. Hughes noticed that students still had trouble formulating introductions and conclusions. She and her students studied and analyzed examples from books and eventually defined additional criteria for introductions and conclusions. These, along with other modifications, were incorporated in a revised assessment form (Figure 2).

A significant component of student self-assessment is having students use the criteria to reflect on their growth, whether by sharing their observations with another student, conferencing with the teacher, or composing a written piece for their portfolios.

Here are some excerpts from the students' reflections that illustrate their thoughts about their improvement over the past year.

- I've improved in creative ways to put my ideas.
- I've learned to make short sentences into long ones.
- I used to write like this: "I'm going to tell you about Erica's future." Now I write like this: "Erica has a future that will sweep you off your feet." I think you can see the difference.
- I learned that using adjectives really sparks up your writing.
- I really liked the assessment sheet because it helped me to know what to do better next time.

FIGURE 1

## SPECIAL PERSON

Student Self -Evaluation

Name____________________ Special Person____________________

Date________________

| ✓− | ✓ | ✓+ | A |
|---|---|---|---|
| Incomplete sentences<br>Little or no punctuation<br>Did not indent<br>Less than 1/2 page<br>Messy paper<br>Simple words misspelled<br>Not proofread<br>Doesn't make sense | Complete sentences<br>Appropriate punctuation<br>Indented paragraphs<br>1/2 a page<br>Simple words correctly spelled<br>Neat<br>Topic Sentence | Introduction<br>Conclusion<br>Interesting<br>Sentences begin with different words<br>More than 1/2 page<br>Uses adjectives<br>1 or 2 ✚ sentences. | Longer, detailed sentences<br>3 or more ✚ sentences<br>✚ sentences contain:<br>• humor<br>• author's voice<br>• personal opinion<br>• colorful adjectives<br>• similes<br>• own observations |

Comments:________________________________________

________________________________________________

________________________________________________

## SPECIAL PERSON

Teacher -Evaluation

Name____________________ Special Person____________________

Date________________

| ✓− | ✓ | ✓+ | A |
|---|---|---|---|
| Incomplete sentences<br>Little or no punctuation<br>Did not indent<br>Less than 1/2 page<br>Messy paper<br>Simple words misspelled<br>Not proofread<br>Doesn't make sense | Complete sentences<br>Appropriate punctuation<br>Indented paragraphs<br>1/2 a page<br>Simple words correctly spelled<br>Neat<br>Topic Sentence | Introduction<br>Conclusion<br>Interesting<br>Sentences begin with different words<br>More than 1/2 page<br>Uses adjectives<br>1 or 2 ✚ sentences. | Longer, detailed sentences<br>3 or more ✚ sentences<br>✚ sentences contain:<br>• humor<br>• author's voice<br>• personal opinion<br>• colorful adjectives<br>• similes<br>• own observations |

Comments:________________________________________

________________________________________________

________________________________________________

FIGURE 2

# SPECIAL PERSON

Student/Teacher Assessment

Name____________________ Special Person____________________

Date________________

My Goal today is________________________________________

| ✓- | ✓ | ✓+ | A |
|---|---|---|---|
| Incomplete sentences<br>Little or no punctuation<br>Did not indent<br>Less than 1/2 page<br>Messy paper<br>Simple words misspelled<br>Not proofread<br>Doesn't make sense | Complete sentences<br>Appropriate punctuation<br>Indented paragraphs<br>1/2 a page<br>Simple words correctly spelled<br>Neat<br>Topic Sentence | *Introduction<br>*Conclusion<br>Sentences begin with different words<br>11 or 2 ✚ sentences.<br>1 page long<br>Neat | Longer, detailed sentences<br>Very neatly completed<br>3 or more ✚ sentences<br>✚ sentences contain:<br>• humor<br>• author's voice<br>• personal opinion<br>• colorful adjectives<br>• similes<br>• metaphors<br>• own observations<br>• Creative way to put my ideas.<br>• interesting phrases |

Three reasons why my grade should be ______

1.________________________________________

2.________________________________________

3.________________________________________

Mrs. Hughes' Comments:__________________________________

*Introductions may include
- My opinion
- A general statement about the person
- Description of the person
- Something flashy to catch the reader's attention (Something the reader might not know about the person)
- The Topic

*Conclusions may include
- A general summary My Opinion
- A Look Ahead to the Future
- What I Want People to Remember

- I have learned that the harder you try, the higher grade you get.

The collaborative assessment process also allows and encourages parental involvement. Being able to clearly see what is expected of their children empowers parents to become active in helping their children learn.

Peer input is also important. Sometimes students' most probing thinking occurs during an assessment session with one of their own peers. During such a session, I overheard a student ask her friend, "What do you mean by 'better'?" What an exciting level of critical thinking for a 4th grade student to exhibit! Working with assessment criteria has taught her to demand a clear definition of how her paper could be "better" and not settle for generalities.

### Collaborative Assessment: The Results

Not only did every student in Mrs. Hughes's class accomplish the 4th grade writing outcomes, the low-achieving students made the most profound improvement. "The assessment process gave them a clear target to hit," says Mrs. Hughes. "They knew what to do and they did it."

Other educators who have used collaborative assessment uniformly experience genuine excitement at the remarkable success of all their students. They tell me that collaborative assessment has revolutionized their way of thinking about their instruction as well as the importance they place on student input.

Once again, I emphasize that collaborative assessment crosses all grade levels and all subjects. It requires teachers to put considerable effort into developing and articulating clear criteria and working collaboratively with students to assess and advance their learning. I predict, however, that when teachers see the results they will welcome collaborative assessment as an integral part of the classroom learning environment.

To that end, remember the little girl whose writing about Andrea was quoted earlier? This is the way she wrote about Marty only three months later:

> Marty is one of the twenty-eight students in our fabulous class. Now lets think way back when Marty was a short-snort.
>
> In pre-school Marty went to Grandma's Place. On Marty's first day she cried. Poor Marty! Ellen was Marty's teacher. There was also another teacher that the kids called "Grandma." Marty thought they were amiable.
>
> For kindergarten Marty went to Perry School. That was the same place I went to kindergarten. In kindergarten we got awesome treats. Marty liked all the trips that she went on.
>
> Ardis was where Marty went for first grade. Her teacher's name was Mrs. Heller. Marty's teacher was kind. Gym and music were what Marty liked about her school. There was also a fair Ardis had once a year. How interesting!
>
> During second grade Marty was still going to Ardis. Her teacher's name was Mrs. Simpson. She was nice but strict. Brandy was Marty's friend.
>
> Mrs. Hughes was Marty's outstanding teacher. At Carpenter we get a huge time for lunch period. Here the work is ten X easier. Marty's incredible friends are: Paris, Kate, Gary, Wanjiku and Imani.
>
> Now you know about part of Marty's school history. I hope you enjoyed reading this just as much as I enjoyed writing this. If you have any more questions just ask Marty.

Progress like this is what collaborative assessment is all about.

# IV-12

# Retelling: Reading Assessment That's Also Good Instruction (or, Reading Instruction That's Also Good Assessment)

Jane B. Braunger
Northwest Regional Educational Laboratory

Donald Graves describes portfolios as "simply too good an idea to be limited to an evaluation instrument" (Graves 1992, p. 1). Their primary value, he asserts, is in students' learning, specifically self-reflection, which leads to increased understanding and the production of quality work. I wish to begin with a similar assertion about retellings. Retellings are excellent for reading assessment. They provide "the most straightforward assessment of the result of the text-reader interaction" (Johnston 1983, p. 54). But having said that, I must emphasize that the reason retellings can become authentic, useful assessment is they *begin* as good instruction in the classroom. Think of assessing retellings as a periodic analysis of written or oral work regularly produced as part of instruction. Students develop the ability to produce better retellings as they practice because retellings are grounded in good reading theory and they can actually improve the skill— reading comprehension— which they assess.

## Retellings Complement a Constructivist Model of Learning

Basically, what students do in creating a retelling is to retell, orally or in writing, a selected text so as to convey their understanding of it, including all relevant details, responses, inferences, and associations. A common prompt is to ask students to "retell what they have just read as if they were telling it to another student who had not read the piece."

What makes retellings such powerful tools for developing and assessing students' reading abilities? Brian Cambourne, an Australian literacy educator who has researched retellings as both instruction and assessment, says that retellings give students an opportunity to use language to learn about language

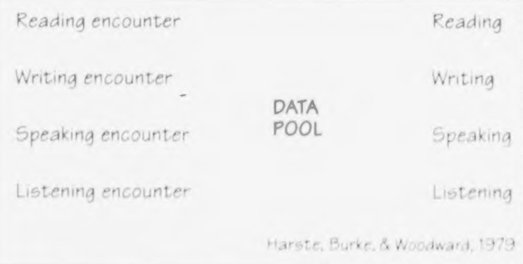
(1987). A retelling is constructed entirely by the student. Teacher- or text-designed comprehension questions stipulate what should be understood and remembered in text. But in a retelling, students use language to frame and express their understanding of the reading. Retellings allow a teacher to discover what students know without prompting them with questions (Rhodes and Shanklin 1993). Retellings mirror the reading process because they are a construction of meaning. They reflect reading as a transaction involving the reader, the text, and possibly other readers; they honor the primacy of the reader's response, and, unlike standardized comprehension measures, they value and account for divergent responses (Tierney, Carter, and Desai 1991). Responses to a text may vary, for example, on the basis of students' experiences and interest in the topic.

Retellings are based on a transactional or constructivist view of literacy, with the following principles:

- Language and literacy are about making meaning.
- Both are constructed through social interaction.
- Both are learned through purposeful use.
- Readers construct meaning through interacting with text and using prior experience and knowledge in the process.
- Learning proceeds from whole to part.
- Reading, writing, speaking, and listening are interdependent and mutually supportive.

This final point is central to the instructional power of retellings, described by Cambourne (1988) as "linguistic spillover," a process in which linguistic forms, ideas, structures, and conventions from the texts used for retellings appear over time in students' reading, writing, and speaking. Cambourne builds on the image of the linguistic data pool (Figure 1). Retellings capitalize on this natural interaction to enhance all language forms.

FIGURE 1
LINGUISTIC DATA POOL



Harste, Burke, & Woodward, 1979

Retellings involve a two-way dynamic. Students pull from the text as they construct a retelling, but they also act on the text, making it their own. Teachers can see instances of students using their own language in retellings to capture the flavor of a piece they've read and communicate their enjoyment of it. Such flexibility across language reflects the students' comprehension through active engagement with the text, as illustrated by the following example.

Some 8th grade students had their first experience with retellings using "Talk: An Ashanti Tale," an African tale about animals, plants, and inanimate objects acquiring the power of talk and using that power to chastise the people who use them (Christ 1989, p. 140). These students composed retellings of varying length and complexity. When they shared their retelling with a partner, and then in a group of four or five, they commented on these differences. They especially enjoyed retellings with a strong voice that incorporated familiar language while remaining true to the original story line. Here is one such example. The original text began like this:

> Once, not far from the city of Accra on the Gulf of Guinea, a country man went out to his garden to dig up some yams to take to market. While he was digging, one of the yams said to him, "Well, at last you're here. You never weeded me, but now you come around with your digging stick. Go away and leave me alone!"
>
> The farmer turned around and looked at his cow in amazement. The cow was chewing her cud and looking at him.
>
> "Did you say something?" he said.
>
> The cow kept on chewing and said nothing, but the man's dog spoke up. "It wasn't the cow who spoke to you," the dog said. "It was the yam. The yam says leave him alone."
>
> The man became angry, because his dog had never talked before, and he didn't like his tone besides. So he took his knife and cut a branch from a palm tree to whip his dog. Just then the palm tree said, "Put that branch down!"
>
> The man was getting very upset about the way things were going, and he started to throw the palm branch away, but the palm branch said, "Man, put me down softly!"
>
> He put the branch down gently on a stone, and the stone said, "Hey, take that thing off me!"
>
> This was enough, and the frightened farmer started to run for his village . . .

Here is one student's written retelling of this part of the text.

> A farmer guy was digging up a yam and the yam got mad at him and told him to leave him alone and the dog got smart with the farmer when he didn't know it was the yam,

> so he was gonna whip the dog with a branch and the tree punked him off and then the branch did too, so he ran away . . .

Retellings provide information on students' reading process and on their understanding of a text. Unlike standardized measures, which cannot account for the reading strategies used in arriving at an answer, retellings allow the teacher to observe the reader's processes, and to hear or read the products of the reader's comprehension. Retellings can be a part of a balanced literacy-assessment program, providing observations of both process and product, as depicted by Anthony, Johnson, Mickelson, and Preece (1991) in their "quad" (Figure 2).

## Benefits as Instruction and Assessment

Both teachers and students can benefit from retellings—students in improved comprehension ability and familiarity with characteristics of various genres, and teachers in deepened insight into students' reading interests, strategies, and understandings.

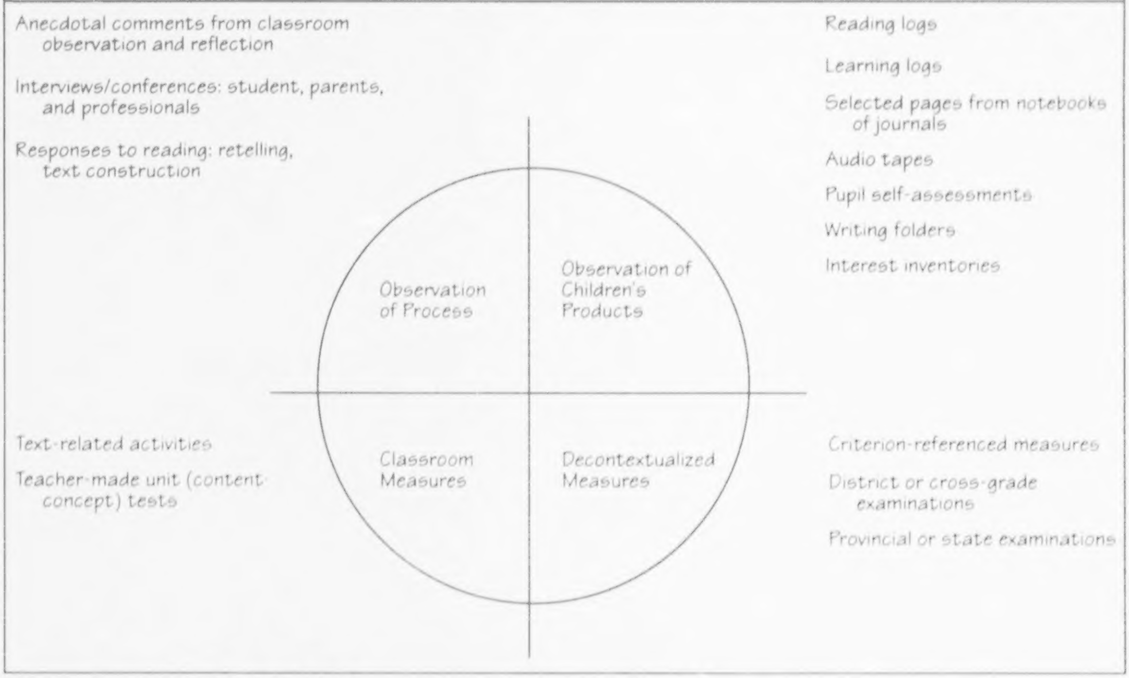
### Benefits to Students

Research with young children has involved asking them to retell fictional or expository material that has been read to them. Benefits include development of more sophisticated language structures, accumulation of background experience, and development of interest in learning to read (Morrow 1988). For example, after reading the book *Tacky the Penguin* to a 1st grade class, I invited them to retell the story. As various children retold favorite parts of the story, I wrote their composite retelling on a chart pack. Predictably, momentum built for the retelling after the first few contributions, and soon the room echoed with shouts of "And then he . . . ," "It was funny when he . . . ," and "But then what happened was . . ." Several sheets of paper and about 20 minutes later, they helped me read their group retelling aloud. It contained all the major events of the story, the faithfully repeated refrain "Tacky was a weird bird," as well as students' responses to characters and situations in the story. The children planned to illustrate their version and bind it so they could enjoy reading it—just like the book it was based on—again and again. The supportive, interactive group process here had the immediate benefit of increasing students' engagement with and enjoyment of the text. It also provided a safe group experience as a model of how students would later produce retellings on their own.

Studies of retelling with older students, 4th graders who included less-proficient and proficient readers, showed improvement across identified com-

FIGURE 2

THE QUAD AS A DATA-GATHERING PROFILE


Anecdotal comments from classroom observation and reflection
Interviews/conferences: student, parents, and professionals
Responses to reading: retelling, text construction
Reading logs
Learning logs
Selected pages from notebooks of journals
Audio tapes
Pupil self-assessments
Writing folders
Interest inventories
Observation of Process
Observation of Children's Products
Classroom Measures
Decontextualized Measures
Text-related activities
Teacher-made unit (content concept) tests
Criterion-referenced measures
District or cross-grade examinations
Provincial or state examinations

BEST COPY AVAILABLE

prehension tasks in both groups after only four practice sessions (Gambrell, Koskinen, and Kapinus 1991). The dimensions of comprehension included setting, theme, plot, and resolution. Students improved their recall of text-based propositions and story-structure elements. The study supported the hypothesis that by orally reconstructing silently read discourse, the reader learns something about organizing and remembering text-acquired information (Gambrell et al. 1991, p. 360). The researchers found that retellings' focus on the whole story supports the students in holistically framing their understanding and encourages elaboration within that framework. In short, active engagement in retelling a story actually improves students' ability to read and understand stories. And comprehension does proceed from whole to part; details have meaning only insofar as they relate to an overall understanding of the text.

Brian Cambourne's work with retellings sheds some light on why this verbal rehearsal technique seems to improve both comprehension and recall of text. Active engagement in meaningful literacy experiences is a powerful tool for learning. The key to student engagement, he says, is in the *immersion* of students in texts of all kinds and the provision of many *demonstrations* of how texts are constructed and used. *Engagement* that accompanies immersion and demonstration occurs when: (1) learners believe they are potential performers or doers of these demonstrated operations, (2) learners see the task as purposeful and useful to them personally, and (3) learners are not afraid that a less-than-perfect performance will be criticized (Cambourne 1988).

The experience of the young child learning to talk is instructive for teachers working with students as beginning or developing readers and writers. Adults continually engage the child in "conversation," clearly expecting the child to enter into the dialogue ever more competently. The child shares this expectation, innately believing, "I can learn to do this." This belief is fundamental to the successful use of retelling. At the heart of retellings' power to improve reading comprehension and spoken and written fluency is the holistic, social-constructivist context in which these activities occur. Students participate in retellings because they learn things to make their own reading and writing better, they are supported by the social context of the classroom as they practice retellings, and they expect to improve over time.

In a typical instructional use of retellings, students may share their written retellings with a partner or read them to the group. The emphasis is on noticing ways in which students' retellings are similar to and different from each other and the text. In this way, students attend closely to language and other text structures, enriching their own "linguistic data pool" in the process. In a supportive social context, students often notice why they prefer another's retelling to their own, and transfer that awareness to their next retelling. Working with a small group of 3rd and 4th graders in a Chapter 1 program, I observed this behavior.

Jenny had been the first to share her retelling of a fable about a tiger and an ant, in which the ant king repays the tiger king for sparing his life by freeing the tiger king, trapped in a cave after an earthquake seals the opening. Jenny's retelling contained detailed accounts of the first meeting of the ant and the tiger, but ended abruptly without including the central event—the tiger being trapped in the cave and rescued by an army of ants. As she listened to Matthew read his retelling next, she smiled and exclaimed, "Oh, that's what I left out. That's what happened." Matthew had written how "everybody tried to help the king: the elephant tried, the monkeys tried, the buffalo tried. The king of the ants came and they (all the ants) took one grain at a time to get the king out. It took a whole day and they freed the king." Matthew had the positive experience of demonstrating a full retelling, and Jenny took useful information from this demonstration. Far from feeling embarrassed or disappointed, she was pleased to get help with making her retellings richer, more complete. Both students also benefited from their extended discussion about the original story, referring to it frequently as they compared their retellings. This type of prolonged encounter with text is another reason that retellings help students improve as readers.

Retellings enrich students' language in all its forms—reading, writing, listening, and speaking. The benefits of classroom use of retellings include greater oral language complexity, improved reading comprehension, increased awareness of the different ways in which texts are structured, more articulation of connections within and between texts, and greater use of literacy language and genre-specific conventions in students' own writing and speaking (Anthony et al. 1991).

### Benefits to Teachers

Retellings provide the teacher with a window into students' reading strategies, background experiences relevant to the text, and understandings of particular texts. Teachers can learn, among other things, what the student thinks is important to remember, whether the student's retelling fits the purposes set for reading, the degree to which the student's responses are text-based or reader-based, and how close the match is between text and retelling (Rhodes and Shanklin 1993, p. 233). An individual retelling may suggest useful literacy demonstrations or experiences for a student. And over time, assessed retellings allow teachers to observe changes in students' reading strategies, attitudes, and comprehension.

In the case of Matthew's and Jenny's retellings, the teacher gained some valuable insights into the students' strengths and needs as readers and writers because the students produced written retellings. Matthew's retelling was quite detailed, with a storyteller's flair for building suspense. Matthew's written retelling, however, revealed some major spelling confusions. Although retellings should not be evaluated for spelling, Matthew's teacher might see spelling patterns in the retelling that, when viewed in the context of writing samples in his folder, would help her to set some realistic spelling targets with Matthew.

Perhaps most important, retelling does not require the teacher to take time away from instruction to gather assessment information. The assessment of retellings is easily accomplished by slightly altering the administration procedure. So, retelling provides useful information to teachers about students' reading and writing development. And students develop into better readers and writers because the retelling process mirrors strategic reading behavior: prediction, sampling, monitoring and revising predictions in light of new textual evidence, questioning, connecting to one's own experience, and summarizing to check understanding.

## Procedures for Eliciting Retellings

Retellings may be done orally or in writing, from text the student has either read or listened to. Obviously, oral retellings are more appropriate until students are able to write independently. As this section will explain, written retellings allow for more student interaction and learning from each other; however, even older students sometimes prefer to give oral retellings if, for example, they are less confident about their writing than their speaking. Students may retell individually or collaborate on a retelling, and either the student or the teacher can choose the text. Whatever the format, students should be told before reading or listening to a text that they will be asked to retell it. And, of course, the use of retellings for assessment should come after students have had extensive experience with retellings in instruction.

The process described here is adapted from the explanation by Anthony and colleagues (1991), based on Brown and Cambourne's model (1987). Additional procedures for taking oral and written retellings are described by Rhodes and Shanklin (1993), Glazer and Brown (1993), and Morrow (1988, 1990). In this procedure, students produce written retellings, interacting with a small group of four or five students:

1. Students see only the title of the text; they write predictions about what they think will be in a text with that title. They can even predict specific words or phrases they'll expect to find.

2. Students share their predictions with their small group. The teacher can then ask for some groups to share with the whole class.

3. Students read the text silently, remembering their original predictions and confirming or modifying them as the text unfolds. Students should be encouraged to read the text as many times as necessary to feel comfortable writing a retelling without referring back to it. They should begin their retelling when they think they understand the text well and can retell it to someone who has not read it. Emphasize that the goal is not to memorize the text, but to retell it in their own words and in their own style.

4. When each student is ready, the text is put aside. Students retell, using appropriate prose, pictures, or diagrams to convey the contents of the text so that someone unfamiliar with it could understand it. If the retelling has a particular purpose, students should be told in advance. For example, students can be asked to retell a piece of historical text to show their understanding of the sequence and relationship of events. Or they can be asked to retell a short fictional piece to show their understanding of the characters' motives. However the purpose is described, students must know that what matters is their ability to convey an understanding of the text, not their mastery of written conventions or penmanship.

5. After students complete their retellings, they share the retelling with a partner. They should be encouraged to note similarities and differences—in the information included, their manner of retelling, etc.—and to refer back to the original text to see more points of comparison and contrast with their retellings. They can also use the original to check omissions and interpretations open to question.

With oral retellings, students may simply be asked to retell a selected text in their own words—for example, as if to a friend who hadn't read it. In an instructional setting, a teacher can prompt the retelling, as needed, asking for more information ("What happened next?"), probing for understanding ("Why do you think he did that?") or seeking conclusions ("What do you think the author was trying to say?"). For students who experience difficulty retelling a story, Rhodes and Shanklin recommend applying the principles of dynamic assessment. Before students read the text, the teacher should provide instruction that's likely to help them construct a good retelling—for example, help students preview the text features and make predictions based on them. Or the teacher could frame the piece, giving students a general introduction to the story and the problem the character in it will face. Then students read and retell the story.

Morrow (1988) reminds us that prompting is instruction, so if a teacher is using a retelling for assessment, he should do so without prompting students because prompts influence students' response to the reading. Glazer and Brown (1993) distinguish between *unguided* and *guided* retellings. In an unguided retelling, students retell the story without any intervention or interaction with peers, teachers, or other listeners. In a guided retelling, prompts, usually posed as questions, help students who experience difficulty moving independently through a retelling.

Retellings used in instruction often include some teacher guidance and, especially with young children, props such as felt boards, puppets, or pictures. The retelling in this instance is a teaching incident. In contrast, retelling used for assessment is generally done without prompts, props, or access to the text upon completion of the reading. The intent here is to gather information about the students' unsupported reconstruction of a text. The value of retelling for assessment depends on the students' familiarity with the process from experiencing it frequently in an instructional context.

## Analyzing, Scoring, Reporting, and Using Retellings

Analysis of a retelling for assessment purposes can be done either qualitatively or quantitatively. Generally, a teacher reads or listens to a retelling to see if a student

- Organizes and sequences ideas clearly
- Uses phrases, vocabulary, and language structures from the original text
- Re-creates the feeling or mood of the original text
- Draws conclusions or inferences from the text
- Summarizes in own words (Ministry of Education and Training 1991).

Recording information on these features can be done anecdotally or on a matrix. Morrow (1988) designed the "retelling profile" to record information about the reader's comprehension of textual information, strategies used, metacognitive awareness, involvement with text, and language facility. The Portland (Oregon) Public Schools (1989) adapted that profile as the Retelling Checklist (Figure 3).

Guidelines for recording information on retellings of narrative pieces are available from numerous sources (Rhodes and Shanklin 1993, Morrow 1990, Glazer and Brown 1993, Cambourne 1988). Common features on which to record information include introduction of the story, setting, plot (sequence of events, problem, resolution), characters (inclusion of significant ones, with adequate detail), theme or point of the story, and use of language or stylistic features from the story (Anthony et al. 1991).

The advantage of a qualitative analysis of retellings is the holistic assessment of comprehension it provides. As Morrow (1988) points out, such analysis is based on the premise that the whole piece is more important than any of its parts, and that the reader's total impression of a text includes those elements. But Morrow also devised a quantitative analysis of retelling, which allows teachers who like the structure of a scorable instrument to account for features of text included in students' retellings. Brown and Glazer (1993) have adapted Morrow's scoring system, and designed a version for student self-assessment (Figure 4).

Figure 3

RETELLING CHECKLIST

Name ____________________

Date ____________________

Title ____________________

| | Low Degree | Moderate Degree | High Degree |
|---|---|---|---|
| Includes information directly in text. | | | |
| Includes inferred information. | | | |
| Includes what is most important. | | | |
| Includes summary or generalization. | | | |
| Includes connections to reader's life. | | | |
| Includes an attachment to the reading (likes or dislikes). | | | |
| Recognizes author's organization and audience. | | | |
| Asks additional questions. | | | |

*Source: Reading Assessment, Portland Public Schools, Portland, Oreg. 1989. Reprinted with permission.*

## Conclusion

Whether qualitatively or quantitatively analyzed, retellings provide rich data on students' reading comprehension, as both a process and a product. If standardized assessment provides a snapshot of students' learning, retellings can be described as a videotape, capturing patterns of growth by revealing

FIGURE 4

## SCORING SYSTEM FOR A NARRATIVE RETELLING

Child's Name ______________________ Age __________ Grade __________

Title/author of Story ______________________ Date __________

Type of retelling:

Circle one from each pair: (Guided/Unguided Oral/Written Listened to/Read)

Directions: Place "1" next to each element, except as indicated. Partial credit can be given only for characters and episodes (e.g., 4 out of 5 characters equals .8). Highest score for retelling is 10.

| SENSE OF STORY STRUCTURE | | COMMENTS |
|---|---|---|
| SETTING: | | |
| Begins story with introduction | ______ | ______________________ |
| Includes time or place | ______ | ______________________ |
| CHARACTERS: | | |
| Names main character | ______ | ______________________ |
| Number of other characters named | ______ | ______________________ |
| Actual number of other characters | ______ | ______________________ |
| Total score for other characters | ______ | ______________________ |
| PROBLEM: | | |
| Refers to main character's goal or problem | ______ | ______________________ |
| EPISODES: | | |
| Number of episodes recalled | ______ | ______________________ |
| Actual number of episodes | ______ | ______________________ |
| Total score for episodes | ______ | ______________________ |
| Proper sequence of episodes | ______ | ______________________ |
| SOLUTION: | | |
| Names problem's resolution | ______ | ______________________ |
| Ends story | ______ | ______________________ |
| THEME: | | |
| States theme of story | ______ | ______________________ |
| Total score for retelling | ______ (out of 10) | |

Affective (personal) involvement with text:

Summary:

Teacher: ______________________

Scoring system originated by L.M. Morrow in "Retelling Stories as a Diagnostic Tool," from Reexamining Reading Diagnosis: New Trends and Procedures, Susan M. Glazer, Lyndon W. Searfoss, and Lance Gentile, eds., copyright 1988, The International Reading Association. Reprinted with permission.

BEST COPY AVAILABLE

the strategies in use over time (Anthony et al., p. 9). Furthermore, retellings are assessment embedded in the learning experience. Students can use what they learn from their retellings to become better readers and writers. Figure 5 summarizes the benefits of retellings in contrast to traditional comprehension tests.

FIGURE 5

### RETELLINGS VS. COMPREHENSION TESTS

| **Retelling** | **Comprehension Test** |
|---|---|
| Student organizes format and determines content; can be a conversation | Test or teacher determines content and format; can be a "gentle inquisition" |
| Involves active construction of meaning | Calls for passive response to questions |
| Provides evidence of student's reading processes and strategies | Does not indicate processes and strategies known and used |
| Indicates prior knowledge | Cannot account for prior knowledge |
| Indicates interest, creative engagement | Does not assess interest, enjoyment |
| Can be a supported effort, via prompts, collaboration | Demands one-shot, perfected response |
| Actually improves comprehension | Takes time from real reading and instruction |
| Values divergent, creative thinking | Prompts convergent thinking |

## REFERENCES

Anthony, R. J., T. D. Johnson, N. I. Mickelson, and A. Preece. (1991). *Evaluating Literacy: A Perspective for Change.* Portsmouth, N.H.: Heinemann Educational Books, Inc.

Brown, C. S., and S. M. Glazer. (1993). *Portfolios and Beyond: Collaborative Assessment in Reading and Writing.* Norwood, Mass.: Christopher-Gordon Publishers, Inc.

Brown, H., and B. Cambourne. (1987). *Read and Retell.* Portsmouth, N.H.: Heinemann Educational Books, Inc.

Cambourne, B. (1988). *The Whole Story: Natural Learning and the Acquisition of Literacy in the Classroom.* Auckland, New Zealand: Ashton Scholastic Limited.

Christ, H. (1989). "Talk, An Ashanti Tale." *In Myths and Folklore,* pp. 140–142. New York: Amsco School Publications.

Gambrell, L. B., B. A. Kapinus, and P. S. Koskinen. (1991). "Retelling and the Reading Comprehension of Proficient and Less-Proficient Readers. *The Journal of Educational Research* 84, 6: 356-362.

Graves, D. H., and B. S. Sunstein, eds. (1992). *Portfolio Portraits.* Portsmouth, N.H.: Heinemann Educational Books, Inc.

Harp, B. (Ed.) (1991). Assessment and Evaluation in Whole Language Programs. Norwood, Mass.: Christopher-Gordon.

Harste, J.C., C.D. Burke, and V.A. Woodward. (1983). "Children's Initial Encounters with Print." NIE Grant Proposal, 1979. Cited in V.H. Hardt,ed. *Teaching Reading with the Other Language Arts.* Newark, Del.: International Reading Association.

Irwin, P. A., and J. N. Mitchell. (in press). *The Reader Retelling Profile: Using Retellings to Make Instructional Decisions.*

Johnston, P. H. (1983). *Reading Comprehension Assessment: A Cognitive Basis.* Newark, Del.: IRA.

Lobel, A. (1980). "The Elephant and His Son." In *Fables,* p. 32. New York: HarperCollins Children's Books.

Ministry of Education and Training. (1991). *English Profiles Handbook: Assessing and Reporting Students' Progress in English.* Victoria, Australia: Ministry of Education and Training.

Morrow, L. M. (1985). "Retelling Stories: A Strategy for Improving Young Children's Comprehension, Concept of Story Structure, and Oral Language Complexity." *Elementary School Journal* 85, 5: 647-661.

Morrow, L. M. (1988). "Retelling Stories as a Diagnostic Tool." In S. M. Glazer, L. W. Searfoss, and L. Gentile, eds. *Re-Examining Reading Diagnosis.* Newark, Del.: IRA.

Morrow, L. M. (1990). "Assessing Children's Understanding of Story Through Their Construction and Reconstruction of Narrative." In *Assessment for Instruction in Early Literacy,* edited by L. M. Morrow, and J. K. Smith, pp. 110–134. Englewood Cliffs, N.J.: Prentice Hall.

Portland Public Schools.(1989). *Reading Assessment: Recording Student Progress.* Portland, Oreg.: Portland Public Schools.

Rhodes, L. K., and N. Shanklin. (1993). *Windows into Literacy: Assessing Learners K-8.* Portsmouth, N.H.: Heinemann Educational Books, Inc.

Tierney, R. J., M. A. Carter, and L. E. Desai. (1991). *Portfolio Assessment in the Reading-Writing Classroom.* Norwood, Mass.: Christopher-Gordon Publishers.

BEST COPY AVAILABLE

# IV-13

# Portfolio Reflections in Middle and Secondary School Classrooms

ROBERTA CAMP
PHILADELPHIA, PENNSYLVANIA

Reprinted from *Portfolios in the Writing Classroom: An Introduction,* edited by K. Yancey (Urbana, Ill.: National Council of Teachers of English, 1992), pp. 61–79. Copyright © 1992 by the National Council of Teachers of English. Reprinted with permission.

# Portfolio Reflections in Middle and Secondary School Classrooms

**Roberta Camp**
Educational Testing Service
Princeton, New Jersey

In writing as in other performances, we learn in part by looking back on what we have done. In this sense, looking back—reflecting—on the experience of writing a piece or on the written piece itself is an integral part of our becoming more accomplished writers. Yet the teaching of writing as conducted in many of our schools and classrooms has not often included opportunities for student writers to reflect on their work.

All too frequently, as Brian Johnston has pointed out in his insightful and helpful book (1983, 1-4), we have interrupted the cycle that is necessary for effective learning. At the point in the cycle where students need to reflect on the experience of writing and articulate what they have learned from it, we have traditionally taken from them both the opportunity and the responsibility for reflection by doing it for them. As teachers responding to student work or as outside evaluators, we have told students what *we* see and what *we* value—very often at precisely the point in their learning where they should be discovering what *they* see and what *they* value. In doing so, we have reduced the likelihood that students will use past experiences in writing to shape subsequent experiences. We have eliminated valuable opportunities for students to learn about themselves as writers—and for us to learn about them.

Fortunately, with the introduction of portfolios into our classrooms, we can provide students with opportunities for reflection that have not previously been available. At the very least, as a mere repository for a student's writing, a portfolio is a record of past writing experiences in

---

The work described in this chapter has been supported by the Rockefeller Foundation.

the form of the products resulting from the experiences. Having access to the record makes it possible for both teacher and student to look back at the writing. But a portfolio can be far more than a repository or record. How much more depends in large part on what teachers and students do with the writing they collect for the portfolios and on the roles they assume in relation to the writing—on the procedures they use for looking at writing, what they attend to, how they talk with one another about what they see.

As portfolio programs are introduced into large numbers of classrooms and schools, they will necessarily vary with the student populations, the programs of instruction, and the purposes and views of learning they are designed to serve. Some approaches will value reflection more than others, and the activities used to stimulate reflection will differ from one portfolio program to another. By noting the differences among approaches to portfolios, and to reflection in particular, I believe, and by observing the effects of these approaches on students and teachers, we can enrich our understanding of the potential for learning created by portfolios.

The pages that follow describe an approach to portfolios in which reflection is central. In this approach, the activity of reflecting on and selecting from student work becomes the very process by which portfolios are generated, the driving force behind the creation of portfolios. Like other portfolio approaches, this one arises from a particular environment—in this case, one in which the primary purpose for portfolios is to promote learning, the purpose of assessment is to serve learning, and reflection is seen as essential to learning.

## Background: The Environment for Portfolio Development

In the fall of 1986, teachers of writing from several middle schools and high schools in the Pittsburgh School District, the supervisors for language arts and English in the district, and researchers from Harvard Project Zero and Educational Testing Service began to meet monthly to talk about instruction and assessment in writing. We were brought together by Arts PROPEL, a Rockefeller-funded effort to develop assessment closely related to instruction in music, visual arts, and imaginative writing in Grades 6 through 12 in the Pittsburgh public schools.[1]

Teachers, supervisors, and researchers worked together as a team exploring ways of opening up to students the opportunity to engage directly with imaginative writing, particularly poetry and drama. We worked in a cycle of development: the team first designed possible activities, then teachers tried them out with the students in their classrooms, then the team discussed the teachers' experience and the resulting student writing, thereby stimulating a new round of design, tryout, and evaluation. The team first focused on developing sequences of classroom activities that invited students to become perceptive readers and engaged writers of poems and dialogues. As part of these activities, the students began to look back at their own work and reflect on both the choices they had made as writers and the effect of those choices on the pieces written.

The result of our work with these classroom activities was twofold: it was the beginning of an infusion of imaginative writing into the district curriculum and—more to the point here—it prepared the way for portfolios. Through the experience of working together as a team, we learned to think collectively about the relationship of assessment to instruction, a valuable preparation for our exploration of portfolios. Furthermore, the classroom projects themselves fostered a climate conducive to portfolios. To be more specific, they emphasized learning and discovery, making choices and taking risks. They gave students experience with long-term projects requiring sustained interest and effort. They promoted a workshop atmosphere in which students took on an active role in furthering their own learning. And—perhaps most important—they involved students in looking back at their writing and the processes they used in creating it.

It had been clear from the beginning of Arts PROPEL that the team's work would include portfolios. The researchers saw portfolios as a necessary part of the project and a timely response to the changes evident more generally in research on writing and writing instruction. Pittsburgh supervisors looked to portfolios as a natural development for secondary school English and language arts, because folders of student writing were already being collected and not much used in the day-to-day life of the classroom. But it was after nearly two years of designing, trying out, and refining the classroom poetry and drama sequence and intermittently talking about portfolios that we found we were all—teachers, supervisors, and researchers—developing a sense of what portfolios could do for students in Arts PROPEL classrooms. We then knew we were ready for systematic exploration of possible approaches to portfolios.

The third year of the project involved four strands of portfolio exploration. Individual teachers in a few classrooms experimented with a handful of portfolio-related or portfolio-generating activities to find out what might be feasible. Meanwhile, other teachers expanded upon and refined the reflective activities within the classroom projects. The researchers interviewed teachers and students in a number of classrooms to discover more about what students see and value in their writing and what they might be able to see with help from others. Then, at our regular meetings throughout the year, the entire Arts PROPEL team discussed what we had learned so far about reflection and other portfolio-related activities and what we—especially the teachers—wanted portfolios to accomplish.

At the end of the third year we had arrived at an approach to portfolios grounded in what students can see and what they value in their writing and designed to help students assume gradually increased responsibility for their learning about writing. We had begun to realize that what students could learn on the way to creating the portfolios was as important to us as the portfolios themselves. As a result, we thought of portfolios as far more than mere repositories for writing. Arts PROPEL portfolios, we decided, would be generated through a yearlong process in which students would keep all of their writing in folders, then engage in a series of reflective activities. The result of these activities would be the portfolio itself, a smaller selection of writings and reflections providing a portrait of the student as a developing writer.

In the portfolio work of the following year, the fourth year, we further experimented with and refined the activities leading to the creation of portfolios, now observing and discussing the effects of the portfolio-related activities on classroom climate and student learning. During this time, we also looked closely at students' writing folders and portfolios to discover what evidence of students' ability and learning we could see and how we might need to restructure the portfolio activities to yield richer and more interpretable information.

## The Arts PROPEL Model: A Portfolio Built on Reflection

From our experience with the poetry and drama sequences and from our experiments with possible portfolio activities, we had learned that the opportunity to look back at their writing was valuable to students' development as writers and to teachers' understanding of their development. Both teachers and researchers had noticed that the reflective activities helped students become aware of strategies and processes they had used in writing. In addition, the activities encouraged the students to develop criteria and standards for their work. But we also knew that the practice of looking back was neither familiar to students nor easy for them. Recognizing that most students have little experience evaluating their own or others' writing, even in the best classrooms, we were not terribly surprised that they came into the school year expecting that the teacher would tell them what was good or not so good in their work, and that their job would be to understand and accept the teacher's judgment.

It seemed to us, therefore, after four years of experimenting and refining, that the first steps toward creating a portfolio would need to engage students in simple and nonthreatening forms of reflection while providing ample support and building a climate of trust. The early stages would involve teacher modeling and oral reflection. Then, as students showed signs of readiness—ownership of their writing, willingness to engage in dialogue about writing—they could be asked to write their reflections on single pieces. Only after they had engaged with these various forms of reflection and with relatively simple forms of selection did we feel they would be prepared to shape the final version of their portfolios.

### Introducing Students to Reflection: Modeling and Oral Reflection

In the Arts PROPEL portfolio activities as they have evolved so far, the first weeks of the school year are devoted to getting students accustomed to the idea of reflection. Students are introduced to the practice of keeping their writing in folders—all of their writing, including notes and drafts—and to the practice of looking at their own writing, reading it aloud, and listening to others read from their writing. Much of this oral reflection activity in the first weeks and months involves the teacher in modeling questions that writers and listeners can ask of one another, as well as possible responses that each might give in return.

Questions like "What did you like best about the piece?" or "What in the piece would you like to know more about?" or "What did you most want your reader to get from this piece?" help direct students toward content rather than surface features and toward specific aspects of the writing rather than general observations about the piece or its writer. The circumstances of responding to pieces read aloud rather than looking at written work also encourage students to attend to ideas and their development rather than to mechanics or other surface features. In time, students become relatively confident that they can respond to one another's writing in ways that are informative and thought provoking but seldom if ever hurtful. They are then encouraged to carry their experience with questions and responses into interactions with partners or small groups.

During the time students are practicing oral reflection, the teacher's written responses to the students' writing reinforce the messages they are getting from the classroom interactions. On a sheet of paper attached to the piece of writing, the teacher typically offers comments on only two points: one thing that is done well in the writing, and one thing to focus on in future writing. The point of the response is not to tell the student everything that has been done poorly, or even everything that is done well. By now, the student is beginning to understand that many things can be said about any piece of writing and that no one person—not even a teacher—is necessarily going to say it all. The point is rather to recognize successes, strengths that can be built upon, and to think about next steps that are likely to yield a return for effort expended. The same two points of focus—a strength and a goal—can be incorporated in teachers' conferences with students, both unstructured, informal conferences and more formal ones.

Sometime in about October or November, the teachers look for signs that students have become accustomed to the idea of responding orally to one another's writing and are ready to move on to more demanding forms of reflection. In particular, teachers notice whether students have begun to internalize the focus on strengths and goals, rather than error and correctness, as a way to evaluate their progress in learning to write. They also look at both classroom interactions and students' responses to peer and teacher comments to determine whether students are beginning to see themselves as writers: whether they use vocabulary that indicates ownership over their writing, whether they refer to what they have done in previous pieces and what they might do in future pieces, whether they are willing to redraft a piece of writing that is not working well for them, whether they allow more time for writing assignments because they feel a need to share their work with others and to revise.

In making their decisions about students' assimilation of the first, simple forms of reflection and their readiness to move on to more complex forms, teachers are making a professional judgment based on their observations of students and their awareness of students' responses to oral and written comments on their work. This decision is the first major event in contextualized portfolio assessment—assessment grounded in student reflection and exercised in the direct service of instruction and learning.

## The First Written Reflection

When students show these signs of readiness, the Arts PROPEL teachers introduce them to the experience of using writing itself to reflect on their work. Students engage in their first written reflection by answering two, possibly three, questions about a single piece of their writing. The questions echo those they have been using in their oral reflection and the points of focus in the teachers' responses to their writing, but they are somewhat more evaluative:

> What do you like best about this piece of writing?
>
> Which of your writing skills or ideas are you least satisfied with in this piece? Why?

The students answer these questions on a sheet of paper separate from the piece of writing, and they answer them before they receive any written response from the teacher.

The answers give the teacher information about the students' perceptions of their own work, perceptions that the teacher can then take into account in responding to each student's piece. The students' answers to the "Why" portion of the second question are especially valuable because they allow the teacher insight into the students' criteria for writing and into the ways that they apply them to their own writing. In a sense, a student's answers to these questions become the beginning of a shared process of assessment—a dialogue about what the writer, the teacher, and sometimes other students see and value first in a single piece of writing and eventually in multiple pieces of writing.

The sense of evaluation as dialogue between student and teacher is enhanced by activities used to introduce the questions. For example, the teacher may first present the questions accompanied by a piece of student writing from another year or school. Students are asked to generate and then share the answers they would give to the questions, as applied to the sample piece. The teacher encourages students to bring forward a wide variety of answers, then points out the legitimacy of many perspectives in responding to writing. Finally, the teacher presents the answers given by the writer of the piece, explaining that the writer's perspective can be one among many. Students and teacher thus begin to form a community of writers who collectively negotiate the meaning of a piece of writing.

The practice of reflecting on individual pieces of writing is continued for all major writings, both to provide students with ample opportunity to internalize the practice of reflection and to create a frame of reference for the writing when they look back at it later in the year.

Sometimes the same two questions or variations on them are used, sometimes a question specific to the project or assignment is added, or a question that invites students to say what was most important to them in writing a piece.

The questions are not conceptually difficult, but they require a candor that may be challenging to the students, even when they have used them before. If we look at the responses of an eighth-grade student reflecting on an essay she wrote, we find in response to the first question, about what she likes best: "I believe I stick to the topic, and have an interesting way of doing so." For the second question, on skills or ideas she is least satisfied with, we observe this somewhat reluctant response: "maybe show more proof or details." And for the third, on the reason for this perception, we discover a still more reluctant response: "Am not quite proving my point as far as I could go." For this student as for many, the idea that even good writers need to revise, or—more pointedly—that anything in one's own writing might leave room for improvement, is difficult to accept. When she does accept it, however, she opens up new possibilities for learning.

Sometimes the teacher will ask students to respond to one another's writing by using the same questions the writer has used in her own reflections. Usually the teacher requires that each piece of writing receive two such peer evaluations. In these cases, the teacher may become the fourth reader to respond to the paper—after the writer herself, and after the two other students. This practice further emphasizes the sense of community of readers and writers engaging in a sustained dialogue about writing. The fact that the teacher uses the same questions as do the other readers of the paper gives the questions credibility and suggests that the peer evaluations can take their place alongside the teacher's response as part of the classroom conversation about writing.

Students' answers to the questions for a particular assignment are sometimes collected and presented to them as an illustration of the range of perceived "best points" and perceived opportunities for improvement. In the eighth-grade class of the student quoted earlier, for example, the "What do you like best" question evoked responses that were quite varied. For example:

> I like the length of my paper. I sometimes write too much or too little.
>
> I like the particular examples that I cited, which may not have been cited by other writers.

> I like how I was able to change my paper so much from the rough draft.

In answer to the "least satisfied" question, students' responses were equally varied:

> I think the way the piece flows is the skill I'm least satisfied with. I think it could have been more smoothly written.
>
> The arrangement of ideas and what I put down wasn't exactly what I thought in my head. It takes away from the writing and what I'm trying to prove, because I assume people know what I'm talking about.
>
> I really didn't have a way to conclude this piece because I couldn't find the right words that would make sense.

From the range of their peers' responses, students can begin to intuit possibilities for looking at writing that they had not themselves envisioned. Their answers to the "Why" question in particular can become the basis for a discussion about the variety of criteria that they have identified collectively and can now use in thinking about other pieces of writing. This process of identifying criteria for good writing is enhanced by posting on blackboard or bulletin board the language that students have used in describing the strengths of their own or others' writing. The posted language then becomes part of the shared discourse of the classroom, a vocabulary enabling students to share in the process of assessment.

### The Students' First Written Perceptions of Themselves as Writers: The Writing Inventory

When students have begun to accumulate a body of work in their writing folders, they are encouraged to use writing to think about and describe themselves as writers. The teacher introduces this form of reflection by showing how she herself would apply a series of questions such as these to her own writing:

> What kinds of writing have you done in the past?
>
> What do you like to do most in writing?
>
> What do you like to do least?
>
> Where do you get your ideas for writing?
>
> What do you think is important to know about you as a writer?

The teacher might also use examples of answers to the questions given by students from previous years or from other schools. The students then answer the questions themselves, using their writing folders to remind them of past experiences with writing. Their own responses to the questions help students become aware of what they value in writing, what resources they bring to it, and where they sometimes struggle with it. Students' answers may again be collated and used as the basis for comparison and discussion, thus encouraging students to see the diversity of their preferences and of their sources of ideas for writing, and possibly suggesting alternatives they had not considered.

The students' responses to the questions also help the teacher discover much that is not otherwise apparent about their values, experiences, and resources for writing. In response to the question about what he likes to do most, one eighth grader said that he likes to write "descriptive yet funny paragraphs" and that "occasionally I write about serious topics. . . . I enjoy writing narratives." Asked what he likes to do least, the same student said, "I dislike writing explanatory paragraphs and research papers." For the last question, about what is important to know about him as a writer, he brought out the distinction between ideas and dislikes more clearly, indicating that he enjoys writing "easygoing paragraphs" but is not good at writing in-depth paragraphs. The teacher might not be entirely surprised by the student's observations, because the narrative pieces in his folder are more accomplished and show fewer signs of struggle than the explanatory pieces. Indeed, his struggle is experienced by many middle school writers who are challenged to move from narrative to expository writing. But his description of it in his own terms gives the teacher important information about the state of his awareness and about the struggle as he experiences it in his own writing.

The information made available to the teacher through the students' reflections is even more valuable when it reveals less predictable resources and struggles. One tenth-grade student, for example, a student for whom writing is often difficult, indicated in his reflections that brainstorming was the process in writing that he liked most because for him "it's the easiest and most interesting"—an unusual view, since many students find the greatest obstacle in the initial idea-generating phases of writing. He liked proofreading least, he said, because "it takes too long"—and it does take long in his case. The most revealing information, however, came in the student's description of what the teacher should know about him as a writer: "I'll sometimes put a whole variety of ideas at once and take forever to choose one. Once I get an idea I tend to stick with it." This is assessment information that is quite useful to the teacher. "Diagnostic" in the best sense, it allows genuine insight into the student's use of processes and strategies for writing, identifies where he has resources and where he experiences obstacles in writing, and indicates his state of awareness about both his talents and the challenges he faces.

### The First Portfolio Selection: An "Important Piece of Writing"

With sufficient opportunity to reflect on individual pieces of writing and the stimulus of a supportive classroom environment, students in time develop the expectation that they will look back at their writing. Because Arts PROPEL students have easy access to the writing folders where they collect their drafts and finished pieces, they eventually begin to refer with increasing frequency to pieces of writing other than the one they are currently working on. Depending on the intensity of students' experience with writing and reflection, as well as other factors related to learning, this change in awareness may become evident after only a few weeks for high school students or after a few months for middle school students. When it does occur, students are likely to have enough experience in looking at their work and a sufficient body of work to take their first step toward selecting from their folder the writing that will go into the portfolio.

In the Arts PROPEL approach, this first selection is quite open to the students' judgment. At this point in the sequence of reflective activities, students have engaged in a number of discussions about their own, their peers', probably even their teacher's writing. In these discussions they have entertained questions from various perspectives about qualities to be valued in a piece of writing. They have seen and most likely have contributed to the development of a classroom list of qualities for good writing. Now they are asked to exercise their own judgment, drawing on their own evolving set of criteria for writing.

Individual students are asked to select a piece of writing that is important to them and to explain why they have chosen that piece—in effect, to explain why it is important to them. The students are also asked to respond individually to a series of questions enlarging the range of their perceptions about the piece. We sense how these questions work if we look at them in combination with a sample of responses from even a few students. Here are the questions, followed in each case by responses from the same three students appearing in the same order:

**Why did you select this particular piece of writing?**

"I believe it's my best piece all year. I think it's a very strong piece."

"It's the most thoughtful piece I have written all year."

"I had to use more references to do this writing, and you can see this by how much more details are in it."

**What do you see as the special strengths of this paper?**

"It shows that I can write a unique piece, different from the rest of the crowd."

"The wording and the form."

"I sense a strong ability to spot details from the text."

**What was especially important when you were writing this piece?**

"I wanted to write something that would stand out, that people would notice. And it was."

"What I thought friendship was all about."

"My main goal was to defend a thesis with as much information as possible."

**What have you learned about writing from your work on this piece?**

"I can begin to write something, and end up with something totally different."

"Writing a poem wasn't as hard as it seems."

"I have learned that when you are writing you must always stick to the topic."

**If you could go on working on this piece, what would you do?**

"I would make it longer, taking off the end, making many more levels of anticipation."

"Be more descriptive."

"I would go into the different ways each of the boys handled their tribes."

**What kind of writing would you like to do in the future?**

"Short stories, POEMS!"

"Narrative."

"I have always wanted to write *a murder mystery*."

First of all, the responses are quite varied, even within a single grade level. Secondly, they show students taking responsibility for their selections and using their own criteria—which vary considerably from student to student—in making their selections. What is more interesting, however, is that they suggest that the students individually are beginning to develop a sense of long-term purpose for writing.

The practice of allowing students to make their own selections for the portfolio, and to do so on the basis of their own criteria, is admittedly risky. It requires sustained and thoughtful preparation, and it makes for messy data—portfolios that do not lend themselves to easy categories for evaluation. Nevertheless, it brings significant rewards for student learning: students' ownership of their writing, of the portfolio, of the values they express in making their portfolio selections.

### The Second Portfolio Selection: Portfolio Update

Several weeks after students have made their first selections for the portfolio, when their interactions around writing and their written reflections show an increased sense of ownership over their writing and increased confidence in their evaluations of writing, they are asked to select a second set of writings. This time the selection process is more challenging: the students are first asked to select a piece they consider satisfying and one they would describe as unsatisfying. They then say why they describe the first as satisfying and what they learned about themselves as writers from their work and reflection on the piece. For the second piece they are asked to indicate with specific reasons why they consider the piece unsatisfying and how they would revise it if given the opportunity.

Having students select a piece that shows them at something less than their best would no doubt be incompatible with many approaches to portfolios, but in Arts PROPEL classrooms the awareness developed through this selection has been valuable to students' learning. In the suppportive environment established around portfolios by March or April, the opportunity to identify an unsatisfying piece from the several months' work in their folders enables students to discover what they have learned about writing, to see how far they have come, and to acknowledge that some pieces of writing are more successful than others, at least in one respect or another. Selected in combination with the satisfying piece, the unsatisfying piece helps students move beyond the view that one is either a good writer or a poor writer, "good at English" or not. The process challenges them to reflect on what Johnston has called "the *specifics* of their experience" and to break down the kind of global evaluations in which "weaknesses are described as if they apply generally rather than in specific definable situations" and "are located squarely within the self" rather than in the writing (1983, 5–7).

In our experience in Arts PROPEL, students' selections on this occasion show that they are developing their own standards for writing. In many cases, students select as "unsatisfying" a piece they earlier considered one of their best works. With increased experience in writing and reflection, these students have discovered that a piece done some weeks or months ago does not compare well with their current writing, or that their standards are now different from those they held earlier. Other students select a piece that does not quite succeed in some respect that is important for them, even though it received a favorable evaluation from teacher or peers. The following three examples, the first two from twelfth graders and one from an eighth grader, illustrate students' reasons for describing a piece as unsatisfying:

> My grade was fine but I don't feel that I applied myself like I could have.
>
> I didn't have enough examples and the ones that I did have I didn't express well enough. This was a hard one to write. I didn't enjoy writing this one.
>
> As I said before, I love describing situations and am very good at it. But in this piece I don't think my skill is all that apparent. Most people think this is one of my best, but I disagree with that. I like my idea, but I think I could express it better.

### Students' Reflections on Processes They Use in Writing: The "Biography of a Work"

Having collected a considerable body of writing in their folders and engaged in reflection on numerous occasions, many students—especially high school students—are able in the later months of the school year to reflect on the processes they have used in creating a piece of writing. These students are asked to select a finished piece, along with notes and whatever other writing they have done in generating it, that will illustrate how they go about the process of writing. The students are asked to respond to a series of questions probing various aspects of the processes and strategies they use in writing, how they felt about the piece while they were writing it, and how the process they used in this piece compares with processes used for other pieces of writing.

Students then think about their answers to the questions and on the basis of their answers write up the "story" of how the work was created. This "story"—the "biography of a work"—is then attached to the piece of writing and accompanied by all the notes and drafts that belong to it. Both of the biographies that follow are from a tenth-grade class.

> **The Biography of a Work**
>
> My writing process begins when I think of a good topic sentence for my paper. After my first sentence is written, my pen begins to flow with ideas and words I never knew existed. Everything is written down like one big mess on paper. The ideas are [then] arranged in a mannerly fashion so they make sense. Nothing complicated, just simple ideas in chronological order.
>
> After everything is in some kind of order, I reread the organized ideas and turn them into a rough copy of my paper. Now the proofreading stage takes place. I think of what could be omitted and what could be placed in. I work on my vocabulary to make the details of my paper sound more sophisticated. My rough copy is jam-packed with new ideas and details. I rewrite my rough copy to make it look more organized.
>
> I proofread once more and after everything is perfect, start writing my final copy. While writing my final copy, I see what once were just ideas turn into a paper that I can be proud of. Finally, it's finished! I turn in my paper feeling good about it because I know I did it right.
>
> **The Birth of David**
>
> While sitting in my room watching *Some Kind of Wonderful*, a movie I highly recommend, the notion of homework came into my mind. Ten-thirty at night is an awful time to start a new English assignment, but I guess you can call it the time of "David's" conception. About half an hour later, my character David came to life as an attractive, inventive, sensitive young man whose hobby was being perpetually cool. After getting his inside and outs generally on paper, I went to a computer thesaurus to look at some unique words to complement David's personality. That was the easiest thing to do. The hardest, as always, was my own faults of writing—the spelling, sentence structure, punctuation. . . . Maybe they're hard because I consider them so trivial. I worked for about 3 nights on the piece, not for that long though. I thought, to make this piece original, I must put in as much of myself as possible.
>
> At last, when I knew that any more detail, dialogue, and sentences would ruin David's secretive nature (totally planned), I gave it my John Hancock and turned it in.

These biographies and the answers to the process questions on which they are based sometimes reveal aspects of the writer's strategies and habits of writing that would not otherwise be apparent to the teacher. The first of the two examples above, for instance, came from a student who writes rapidly and without visible sign of effort. Without the information provided by the student's reflection, the teacher might not have known how much thought the student put into organizing and revising her papers. Even when the revelations in the biographies are less surprising, however, they provide information that helps the teacher direct students toward more effective strategies and processes for writing. This focus on strategies and processes, which Scardamalia and Bereiter (1985) have described as "procedural facilitation," is valuable to students' becoming more accomplished writers.

### Creating the Final Version of the Portfolio

Near the end of the year, students are asked to look at the pieces in their portfolios and to review the writing in their writing folders. They then decide which additional piece from their folders they would like to add to the portfolio—the "Student's Free Pick." On a cover sheet attached to the piece and to the reflection that already accompanies it, students indicate their reasons for selection. Some choose a piece they consider especially well written; others choose one that represents their greatest investment of effort; others choose a paper that shows a quality in their writing that is not well illustrated by the other pieces in the portfolio.

To help students think about the pieces that now make up their portfolio selections, they are presented with the following questions or variations on them:

> What do you notice about your earlier work?
>
> How do you think your writing has changed?
>
> What do you know now that you did not know before?
>
> At what points did you discover something new about writing?
>
> How do the changes you see in your writing affect the way you see yourself as a writer?
>
> Are there pieces you have changed your mind about—that you liked before, but don't like now, or didn't like before but do like now? If so, which ones? What made you change your mind?
>
> In what ways do you think your reading has influenced your writing?

After students have answered the questions, they write in paragraph form a final reflection indicating how the pieces in the portfolio show evidence that they have grown or changed as writers during the school year. They complete a final table of contents for their portfolio, and another for their folder; the former serves as a guide to the portfolio, and the latter indicates the range of writing from which the portfolio selections have been drawn.

The Arts PROPEL teachers experimenting with the portfolios have developed variations on this final shaping activity, as they have for many of the other portfolio reflections. The first two pieces below are reflections from students in a twelfth-grade class whose teacher brought in writing folders from the students' previous years. The last is the final reflection of a student whose class used the questions and followed the procedures exactly as described above; this student was looking only at writing from the current year.

**Student One**

As I look through my old written work, I notice that I grew up a great deal since 11th grade. My writing is more easily understood now. Last year I used single words. In this year's work I see that I have been using more complex vocabulary. Even my handwriting has improved since last year. My spelling has improved a great deal. My sentences are better written. Although I still use run-ons, I don't use them as often as before. I never realized it, but looking through my folders, I see that I have improved. I think my style of writing is more mature. The things I write about are not petty things. I think I have grown up a lot in the past year.

**Student Two**

In 11th grade, it was always hard for me to write a good introduction. I would usually just go right into detail. I wasn't really sure how to write one. I also noticed that I could have used more descriptive words. I used to misspell a lot of words and thought it was a waste to proofread my work. I realize [now] how important proofreading is. When I proofread, I find a lot of mistakes. It is also easier for me to find a good topic, because I concentrate better on my work. That is about the only changes I can see.

**Student Three**

When I look back at my writing from the beginning of the year I realize that I have changed tremendously as a writer. My earlier work is not as explicit and does not seem like anything I would write now. I have found that this year has made the greatest effect on my writing than any other.

I know now that revising your work adds a great deal to the quality of the piece. If I may quote [my teacher], "Nothing is ever perfect the first time." Each piece of writing we did made me realize more and more things that could make my writing better.

> After these changes have been made I find that I look upon myself as a better and more sophisticated writer. At the beginning of this year I thought my "Lady and the Tiger" piece was the best I could do. When I look at it now I see a lot of places in which I could change it to make it 100% better. The superb teaching I received made me change my mind about most of my writing selections.
>
> The reading I have done this year has helped me to connect some of the styles of famous authors with styles of my own. This in itself has helped my reading progress. I wish to thank [my teacher] for teaching me the fundamentals of writing, for helping me to become a better writer, and for teaching me to respect my own writing.

The writers of these reflections attend to different qualities in their writing, but each has criteria and standards for writing. What is more important, each expresses ownership over the writing and the process of development evident in it, even though each would describe that development in different terms.

## Conclusion

For the students whose reflections have been cited in the pages above, as for many others at all levels of ability in Arts PROPEL classrooms, portfolios have provided occasions to learn about writing as a rich and complex activity and to assume increased responsibility for that learning. The portfolios have done as much as they have for these students in large part because they enabled them, with help from teachers and peers, to look back at their writing and to discover what they see and what they value in it. Portfolios and the yearlong process of generating them have restored these students to the full cycle of learning in which they write, reflect on the experience of writing, articulate what they have learned from it, and go on to new experiences with writing informed by what they have learned.

The approach we have taken in Arts PROPEL portfolios requires considerable attention from teachers and students through the course of the school year, but because it is built around students' selections of and reflections on their writing, it addresses their learning directly. We will continue to work with the reflective activities in Arts PROPEL, refining and enriching them even as we turn the major portion of our efforts toward procedures for evaluating the portfolios. We hope that others designing portfolio programs—teachers, supervisors, and researchers in other schools and districts—will make reflection an integral part of their portfolios, so that the students they serve can also use portfolios to learn about their writing and about their development as writers.

## Note

1. The Arts PROPEL team has drawn on the expertise and hard work of a number of teachers, supervisors, and researchers over the years. The Pittsburgh teachers involved in the project include Marilyn Caldwell, Rita Clark, William Cooper, Jane Zachary Gargaro, Jerome Halpern, Rose Haverlack, Lorraine C. Hoag, Kathryn Howard, Diane Hughes, Annette Scott Jordan, Jean Kabbert, Denyse Littles, Daniel M. Macel, Rose McLaughlin, Arla Muha, Joan Neal, Carolyn Olasewere, Cheryl Parshall, Mary Ann Rehm, Elizabeth Sanford, Mary Ann Strachan, Mary Cullinhan Wise, James Wright, Dan Wyse, and Daniel Zygowski. Supervisors include JoAnn Doran, JoAnne T. Eresh, Alice Turner, and Sylvia Wade. Researchers include Roberta Camp and Drew Gitomer from Educational Testing Service, and Steve Seidel, Dennie Palmer Wolf, and Rieneke Zessoules from Harvard Project Zero.

## References

Johnston, B. 1983. *Assessing English: Helping Students to Reflect on Their Work.* Philadelphia: Open Court Press.

Scardamalia, M., and C. Bereiter. 1985. "Research on Written Composition." In *Handbook of Research on Teaching* (3rd ed.), edited by M. Wittrock, 708–803. New York: Macmillan.

# V

# Designing Performance Assessment Tasks

# V-1

# Designing Performance Assessment Tasks

## Introduction

Judith A. Arter
Northwest Regional Educational Laboratory

### Understanding Performance Tasks

Performance assessments have two essential parts—the performance *task* and the performance *criteria*. Tasks are activities that students engage in, in order to demonstrate what they know and can do. For example, the tasks on the driver's test include things such as parallel parking, driving on one- and two-way streets, driving on the freeway, and backing around a corner. Tasks can be specific exercises or problems (for example, math problems, a specific laboratory experiment, a group project, an essay, or developing a portfolio); or a specific circumstance where information about students will be collected (for example, student cooperation might be observed during the course of planning a class trip).

Performance criteria are the other essential ingredient for a performance assessment. Performance criteria specify what we will look for in student responses to a task in order to determine the quality of performance. For example, performance criteria on the driver's test include things such as knowledge of the law, watching other drivers, not causing congestion, ability to handle the car, and maintaining appropriate speed. In other words, the driving inspector looks for these things while the examinee is performing the driving tasks. In a more educational setting consider writing. The writing assignment is the *task*. Criteria are the basis for judging the quality of the student writing produced in the assignment. For example, the quality of student writing might be judged for ideas, organization, voice, word choice, sentence fluency, and accurate conventions.

Both high-quality tasks and high-quality performance criteria are essential. This section focuses on the performance tasks. Papers in this section (1) discuss the role of tasks in a performance assessment; (2) provide examples of tasks; (3) explore the ways that tasks vary and discuss the advantages and disadvantages of various design options; (4) illustrate how to develop tasks; and (5) discuss signs of task quality. (Section VI in this book section emphasizes performance criteria. Since the two parts of a performance assessment are obviously interrelated, several papers in this section also include a discussion of criteria.)

## Types of Tasks

Papers in this section offer many examples of performance tasks. **Sandra Schnitzer** provides an extended example of a high school biology task; **Judith Arter** shows many task examples in science and mathematics; **Vicki Spandel** discusses writing prompts; and **Gail Goldberg** explores written tasks from many subject areas. **Grant Wiggins** provides examples to illustrate his points concerning quality. **Richard Shavelson and Gail P. Baxter** describe three hands-on science investigations.

These authors make it clear that performance tasks are not all alike. There are almost as many variations as there are individuals and organizations developing assessments. Several papers in this section explore different design options for performance tasks and their relative advantages and disadvantages. **Shavelson and Baxter** discuss different ways that science process skills can be assessed (observation, lab notebook, multiple-choice test, etc.) and their interchangeability (or lack of it). **Arter** points out the following differences in task design:

- The use of manipulatives (versus all paper and pencil).
- A single correct answer versus multiple solutions or approaches.
- The complexity of the task (short answer versus classroom-embedded projects lasting several days).
- Students working in groups or individually.
- The amount of time students have to complete the task.
- How tasks are presented to students and how students are allowed to respond (written, verbal, visual/kinesthetic).
- The degree of student choice.

Most authors in this section agree with **Richard J. Stiggins** that tasks should be designed according to what one wants to assess. Thus, constructed response questions with only one right answer might be the best way to assess, for example, the ability to divide. More open-ended and integrative tasks using materials and equipment might be best to assess complex skills and abilities such as science process skills, problem solving, critical thinking, and the ability to apply information to new situations.

Where authors disagree has to do with what they are willing to call "performance assessment." Some assessment developers are willing to call call the above variations performance assessment. Others, including most of the authors in this section, want to restrict the term "performance assessment" to the more complex and integrative tasks used to assess complex student outcomes. Be careful. If you do not know the orientation of the author, it is easy to become confused when they say things like "*all* performance assessments must have features x, y or z." Many times such statements arise because authors are advocating a particular view or definition of performance assessment.

## How to Develop Tasks

The papers by **Stiggins** and **Stephen K. Hess** provide excellent suggestions for developing tasks. Stiggins discusses the decisions that must be made when designing tasks—whether students will be observed naturally during the course of instruction or whether a separate, structured assessment task will be developed. Hess describes the procedure used in Maryland to involve teachers in the design and pilot testing of tasks. He emphasizes the crucial role of teacher involvement and pilot testing.

## Signs of Quality

The role of the performance task is to elicit the right student performance while avoiding things that might interfere with student performance (extraneous interference). Think about it: If a task doesn't make students think critically, we can't assess critical thinking (maybe students are just regurgitating something they read in a book). We can't assess something that students don't do. We also can't assess a skill if we can't see it, so if the task requires critical thinking, but we have no evidence of the thinking that occurs inside the student's head, we will have trouble assessing critical thinking. Likewise, if one skill masks another, we can't assess the masked skill. Examples are too much reading or writing on a mathematics problem-solving assessment, or role playing in a task designed to assess group collaboration.

Most of the papers in this section (**Stiggins, Wiggins, Shavelson and Baxter, Goldberg, Spandel,** and **Hess**) address issues of quality—what makes a *good* performance task. In other words, how can we ensure that the task actually elicits the correct student performance and avoids interference?

Several authors (**Stiggins, Wiggins,** and **Hess**) discuss the quality of performance tasks in general. To really appreciate these papers, classroom teachers might want to take a minute to think about a time in their own classrooms when an activity used with students worked really well. What made it work well? Every group of teachers with whom I have done this activity has generated a similar list: engaging for students, many ways to proceed, no single correct answer, group work, relates to students' lives, enough time and resources, more than one way to respond, student choice, and hands-on.

Not surprisingly, this is the same list of features that **Wiggins, Stiggins,** and **Hess** propose as being qualities of good performance tasks. Wiggins, for example, lists the following: multiple approaches, styles, and solutions; realistic outcomes; thought provoking and engaging student interest; and meaningful. Hess provides a similar list: multiple response modes, engaging, opportunity for creative solutions, authentic and real world, requires in-depth responses, and taps integrated skills. Stiggins restates a list from Joan Baron: rich, integrative, requires students to access prior knowledge, requires sustained effort over a period of time, student choice, personal meaning to students, challenging, and situated in real-world contexts.

Notice that this list tends to be the opposite of how one might describe short-answer items. These authors are operating from the view that performance assessments should be used to assess the more complex skills and abilities we want to foster in students and, therefore, the tasks must be meaty enough to require students to demonstrate those skills.

In addition to the above general discussions of task quality, other authors in this section focus more directly on specific content areas. **Spandel,** for example, provides an in-depth analysis of writing prompts, and **Goldberg** provides an amazing list of the ways that small changes in the wording of questions can affect what students do.

# V-2

# Designing an Authentic Assessment

SANDRA SCHNITZER
AURORA PUBLIC SCHOOLS

As a high school science teacher, I often posed exam questions that required application of the many facts that students learned. Usually I finished grading the exams feeling quite frustrated. The students didn't seem able to use what they had learned in any novel way or new situation. Why couldn't they do tasks that required them to see relationships, compare, or make decisions?

The answer came to me when I began working with the authentic task assessment model that the Mid-Continent Regional Educational Laboratory (McREL) developed for the Aurora Public Schools. What students had learned was missing a crucial piece. I had not taught them the thinking processes that they needed in order to use the knowledge that they had acquired.

The McREL assessment model identifies 14 complex thinking processes: comparison, classification, structural analysis, supported induction, supported deduction, error analysis, constructing support, extending, decision making, investigation, systems analysis, problem solving, experimental inquiry, and invention. These complex thinking tasks can be integrated into a range of authentic tasks and assessments of performance on those tasks.

The timing of my discovery was opportune. The Aurora school district, like many others across the country, had been looking for ways to restructure the curriculum so that it would present more challenging and engaging tasks to the students. Fortunately, the district was able to provide funds to underwrite time for teachers to plan and work collaboratively. Pairs of science teachers at Gateway High School were released from the classroom for two days to design authentic tasks and assessments in their content areas.

My partner, Bob Legge, and I worked together to design an assessment task that would integrate one of the McREL thinking processes into Bob's advanced biology class. Since the class was ready to begin studying the uses and regulation of biotechnology, we set out to design a task that would allow students to demonstrate their biotechnological knowledge and simultaneously incorporate a complex thinking process.

Decision making seemed to be the most appropriate thinking process to concentrate on, because it is a way to resolve complex issues that have no clear-cut answer, the very kind of task encountered

---

First published in *Educational Leadership* 50, 7 (April 1993): 32–35. Reprinted with permission from the author.

**Bob Legge**, Biology Teacher, Gateway High School, assisted with this article.

in determining the uses, possibilities, and limitations of biotechnological applications.

Bob and I recognized that students should be given opportunities to practice with the form of an assessment before it is used. In other words, they need time to practice the decision-making process without also having to attend to content. Consequently, we had students practice making and using a decision matrix that would determine what to do about a hypothetical American hostage situation in Iran (Iozzi and Bastardo 1990). The students were to advise the President on a course of action that would best resolve a situation in which passengers on an airliner were seized as hostages.

A decision-making task requires students to evaluate a list of alternatives that are scored using carefully weighted criteria. In our practice run, students' choice of alternatives (that is, their decision) was based on the alternatives and criteria that we provided. The alternatives to be considered included military action, economic sanctions, a blockade, doing nothing, giving in, and diplomatic intervention.

The selection criteria included hostages' safety, public opinion, world opinion, economic repercussions, potential for war, and more kidnapping (Iozzi and Bastardo 1990). With teacher guidance, the students determined the weighting factors by deciding which of the criteria were most important to a good decision (the weighting factors are proportions of 100 percent; they add up to 1). Students also give each alternative a score, based on the extent to which each choice possessed each criterion. Each score was then multiplied by the weighting factor to obtain a final total score for each alternative. The alternative with the highest total is the decision. These scores were based on previous student knowledge (but, in later assessments, on their research and reading).

The students then began developing the matrix that would determine the best advice to give the President (see fig. 1). The first decision reached was to do nothing. The students were very surprised.

After much discussion, the students decided that the original weight that they had assigned to the world opinion criterion was not heavy enough. A change in the weight of that criteria resulted in a decision to use military action. The students' alteration illustrates an interesting feature of making decisions by using a matrix, and that is that by changing the weights assigned to the chosen criteria, students can see the immediate effect on the decision.

Once familiar with the process, the students were ready to apply their decision-making model to the content in biology. In preparation for this authentic task, the class read and discussed material on genetic engineering and biotechnology—including the late Judge John Sirica's decision to prohibit experiments that would release engineered organisms into the environment. The issue presented to the students was, who should regulate and monitor biotechnology, society or the scientific community (Iozzi and Bastardo 1990)?

Working in cooperative groups, the students determined the alternatives, the criteria, the weighting factors, and the final decision scores. A decision was generated, and as a group, students shared their decisions orally with the rest of the class.

Now that the students had practiced on two types of decisions, they were ready to begin the authentic assessment. It would allow them to demonstrate their decision-making skills and their ability to integrate and apply these skills in new situations. Students were allowed to choose among several controversial issues in biology. The issues included *in vitro* fertilization, alternatives to animal experimentation, and organ donation.

We provided students with initial background information on each issue (the data were taken from *Taking Sides: Clashing Views on Controversial Bio-Ethical Issues, Levine 1990)*. Students cooperated in gathering further information about their issue. We encouraged them to seek information from a variety of sources.

In the assessment, students were to develop a decision-making matrix using criteria, alternatives, and weighting factors to generate scores based on the information that they had gathered. We also asked the students to prepare a visual representation of their decision. It was to be used as groups presented their issue and decision to the class.

To help students develop their representation, we showed them how to change their decision matrix into a Microsoftworks computer spreadsheet. A student-generated spreadsheet on the issue of organ donation lists the alternatives considered: (1) leave the laws as they are currently written (do nothing), (2) harvest organs only with prior permission, and (3) stop harvesting organs entirely. The criteria for making the decision were: (1) improvement of lives, (2) prevention of deaths, and (3) ethical beliefs.

The students' conclusion was interesting and noteworthy. They decided to let the laws governing the removal of organs stand as they are presently written.

We also instructed the students on the computer procedure for converting spreadsheet information into a graph. Once the spreadsheet and graph data were entered, students were able to print both the spreadsheet and graph on a transparency. The students thus created a professional-looking visual aid to help them lead the class discussion about their issue.

## The Assessment of Decision Making

From the outset of this series of activities, students knew that their performance would be assessed according to a model developed by McREL for each of the 14 complex thinking processes. Each of the processes has assessment criteria, which describe the characteristics of someone who uses the thinking process effectively. We changed McREL's wording about decision makers to make the criteria more specific to the task. Our three assessment criteria ultimately were:

- Did the students' final recommended decision match their chosen criteria?
- Did the students apply accurate and important biological information to the process of decision making?
- Did the students select appropriate and important criteria with which to assess the alternatives?

The McREL assessment model also specifies a four-step rubric. It describes observable characteristics by which student performance can be gauged against each of the assessment criteria (Redding 1992). For example, our final criterion for the process of decision making read, "Did the student select appropriate and important criteria with which to assess the alternatives?" The rubric for this criterion reads:

> *Level 4:* The student clearly and completely identified the criteria by which the alternatives were assessed. The criteria were presented in detail and reflected an unusually thorough understanding and concern for the repercussions of the decision.
>
> *Level 3:* The student clearly identified the alternatives to be assessed. With no significant exceptions, the criteria were appropriate to the alternatives and important to the decision task.
>
> *Level 2:* The student correctly identified the principal criteria by which the alternatives would be assessed. Some criteria might be omitted, or included criteria might not be important factors for consideration or entirely appropriate for the decision.
>
> *Level 1:* The student specified criteria that were not appropriate for the selected alternatives or of importance to the decision.

Our students knew that level 3 is the standard, and all work is expected to reach that quality. Level 4 is exemplary work, and we encouraged students to make that their goal. Both Bob and I assessed the results, but students also used the McREL rubrics to assess their own performances.

## What Is the Difference?

This authentic task and the assessment of it was an excellent way for students to both learn information *and* grasp the relationships among the data. Instead of just collecting and sharing factual information with their classmates, students needed to synthesize what they had learned, integrate it into

FIGURE 1
HOSTAGE DECISION-MAKING MATRIX

| | Selection Criteria and Werighting Factors | | | | | | |
|---|---|---|---|---|---|---|---|
| Alternatives | Hostage Safety .28 | Public Opinion .14 | World Opinion .05 | Economic Repercussions .10 | Potential for War .19 | More Kidnapping .24 | Total |
| **Military Action** a score of X | 1 | 7 | 4 | 1 | 5 | 8 | |
| = | .28 | .98 | .20 | .10 | .95 | 1.92 | 4.43 |
| **Economic Sanctions** a score of X | 4 | 7 | 6 | 2 | 2 | 5 | |
| = | 1.12 | .98 | .3 | .2 | .38 | 1.2 | 4.18 |
| **Blockade** a score of X | 4 | 7 | 6 | 2 | 5 | 5 | |
| = | 1.12 | .98 | .3 | .2 | .95 | 1.25 | 4.80 |
| **Do Nothing** a score of X | 7 | 3 | 5 | 8 | 8 | 5 | |
| = | 1.96 | .42 | .25 | .8 | 1.52 | 1.2 | 6.15 |
| **Give In** a score of X | 9 | 3 | 5 | 8 | 8 | 1 | |
| = | 2.52 | .42 | .25 | .8 | 1.52 | .24 | 5.75 |
| **Intervention** a score of X | 5 | 5 | 5 | 5 | 5 | 5 | |
| = | 1.4 | .7 | .25 | .5 | .95 | 1.2 | 5.00 |

the form of a decision, and then justify that decision to others. The students also learned a great deal about the process of making decisions with no clear-cut answers.

The conversations that took place during cooperative learning were lively. They engaged every student. Furthermore, students took responsibility for some of their own learning, especially when they orchestrated the class discussion.

The preparation for the assessment seemed to work very well. The pre-assessment activities provided students with all the skills and experiences needed to perform successfully in the assessment. Students also knew all of the assessment criteria right from the start, so they had a clear, stationary target to aim at.

Students were able to transfer their learning, too. A few days after the student presentations, a representative from the Aurora Water Department was a guest speaker in the class. She presented material on how water was delivered to the city. Then she asked the students what they thought her real purpose was in visiting the class that day. Confidently, a student answered, "To talk about making decisions about water use." The surprise on the face of the guest speaker was apparent.

The animation of the students was even more unmistakable. Eagerly, they helped the guest speaker generate a list of alternate ways to supply Aurora's growing demand for water and a list of criteria by which to judge the appropriateness of each alternative.

Finally, this assessment procedure exposed students to a new way to use technology, namely creating a high-quality product to use in a classroom presentation. Though many students knew about word processing programs, none were familiar with either the spreadsheet or graphing applications.

In short, we think that authentic tasks and assessments, one such as we have described, will help our students graduate better equipped to handle the complexities and uncertainties of the 21st century.**

## REFERENCES

Iozzi, Louis A. and Peter J. Bastardo. (1990). *Decisions for Today and Tomorrow*. Longmont, Colo.: Sopris West, Inc.

Levine, Carol. (1984). *Taking Sides: Clashing Views on Controversial Bio-Ethical Issues*. Guilford, Conn.: The Dushkin Publishing Group, Inc.

Redding, Nora. (May 1992). "Assessing the Big Outcomes." *Educational Leadership* 49, 8: 49-53.

# V-3

# Alternative Assessment in Science and Mathematics

JUDITH A. ARTER
NORTHWEST REGIONAL EDUCATIONAL LABORATORY

Recently, as part of a collaborative project among several federally funded educational research and development institutions, I reviewed more than 100 alternative assessment devices in science and mathematics.[1] Because we wanted to show a wide variety of approaches to and purposes for alternative assessment, we adopted the following broad definition: "Alternative assessment includes any type of assessment in which students create a response to a question rather than choosing a response from a given list (as for multiple choice, true/false, or matching). Alternative assessments can include short answer questions, essays, performances, oral presentations, demonstrations, exhibitions, and portfolios." The following sections describe what we found.

## Dimensions of Tasks

Performance tasks are the exercises or assignments we give students to do; students' performance of the tasks is what we assess. Tasks vary in many ways, such as the number of right answers, use of manipulatives, required response mode, length and complexity, amount of group work, timing, and amount of student choice. Differences reflect our conceptualization of science and mathematics education and our philosophy of instruction. It is useful to discuss some of the differences so readers can relate the possibilities to their own instructional philosophy and goals for students.

### One Correct Answer vs. More Than One

Many tasks are problems for which the student generates one correct or best answer. Even though tasks may be complicated and require equipment or manipulatives, there is still only one right answer or perhaps a few best answers. All the examples in Figure 1 have one correct or best answer. Other tasks are problems for which there is no single right answer. Consider the tasks in Figure 2.

There are many examples of both types of tasks in the science and mathematics alternative assessments we examined. One approach is not necessarily better than another. Tasks with one right answer

---

[1] The project was a collaboration among the 10 regional educational laboratories and CRESST (Center for Research Evaluation, Standards, and Student Testing at UCLA). The labs and CRESST pooled their assessment collections and expertise to make samples more readily available to people and develop training materials to promote good consumerism of the assessments. For more information see *Improving Science and Mathematics Education/A Toolkit for Professional Developers: Alternative Assessment*. Developed by the Laboratory Network Program (Portland, Oreg.: Northwest Regional Educational Laboratory).

FIGURE 1

### EXAMPLES OF TASKS HAVING ONLY ONE CORRECT ANSWER

A. On a school district test:
145 + 28 + 3767 + 90405 = __________

B. Students are given an order form, the prices for various pieces of clothing (and their associated discounts), and a list of what to order. They fill out the order form and add sales tax.

C. Ask children to do two additions and two subtractions using their counters—for example 4 + 5; 6 + 2; 7 – 5 (Whetton 1991).

D. To see whether students can classify different types of solid waste, students are given a garbage can with specified trash and told to:

"Examine items in the garbage can and sort them into the following categories: paper, metal, food waste, plastic, yard waste, glass, other. Count and record the number of items in each category. List the categories in order from least number of items to greatest number of items. Complete a graph of the results" (Lee and Van Pelt 1991).

E. Students are given experimental equipment and are asked to answer this question:

"Do sow bugs prefer it where it is light or where it is dark?" (Pine, Baxter, and Shavelson 1991; Ruiz-Primo, Baxter, and Shavelson 1993; and Shavelson, Baxter, and Pine 1992)

F. Chris wants to find out which of two spot removers is better. First, he tries spot remover "A" on T-shirts that have fruit stains and chocolate stains. Next, he tries spot remover "B" on jeans that have grass stains and rust stains. He compares the results. What did Chris do wrong that will make it difficult for him to decide which spot remover is better (Kentucky Department of Education 1992)?

FIGURE 2

### EXAMPLES OF TASKS HAVING MORE THAN ONE CORRECT ANSWER

G. "Your class will receive a 30-gallon aquarium. The class will have $25 to spend on fish. You will plan which fish to buy. Use the *Choosing Fish for Your Aquarium* brochure to help you choose the fish. The brochure tells you things you must know about the size of the fish, how much they cost and their special needs. Choose as many different kinds of fish as you can. Then, write a letter to me explaining which fish you chose. In your letter tell me how many of each kind of fish to buy, give the reasons you chose those fish, show that you are not overspending and that the fish will not be too crowded in the aquarium" (New Standards Project).

H. "Estimate the number of beans in the bucket." There are a range of best answers, but the purpose of the task is to see how students go about the task of estimating. Students are given a large bucket of beans, magic markers, a tray and a small cup (Hall 1992).

I. Students produce a mechanized vehicle that performs at least two simultaneous motions in different directions to accomplish a set of actions. For example, a dump truck or an oscillating fan (Bennet 1992).

J. This is the second part to example F given above:

"If you want to help Chris find out which spot remover is better, how will you design an experiment" (Kentucky Department of Education 1992)?

K. This is the final part to example D above:

"What can you do to reduce the amount of solid waste in this garbage can" (Lee and Van Pelt 1991)?

seem to be associated with an attempt to measure specific pieces of information or the ability to apply concepts to new situations, whereas more open-ended tasks seem to be associated with an attempt to measure problem solving, group process, critical thinking, or science process skills.

These observations should not be taken to mean that this is how *you* should design tasks. There are also assessments where students are given tasks with a single correct answer and responses are scored for problem solving (e.g., Sample 1 at the end of this paper). You should have a clear idea of what you want to measure and a well-articulated instructional philosophy. Then, choose or design tasks that fit your instructional philosophy and elicit the right behavior from your students.

## How Students Perform Tasks

There are two major ways students perform tasks—with and without the use of manipulatives or equipment. Using manipulatives can also be called "hands on" learning. Hands-on tasks contrast with assessments that call on students to use paper and pencil to manipulate symbols, draw pictures, calculate, and so on. There are many examples of both types of procedures in the assessments we've seen. Examples of tasks using manipulatives or equipment from Figures 1 and 2 are D (solid waste), E (sow bugs), H (bucket of beans), C (arithmetic), and J (mechanized devices). Examples of tasks that are presented entirely in written form and require only paper and pencil for solution are A (calculations), F (spot remover), and G (aquarium).

Both approaches (with and without the use of equipment and manipulatives) have been used to assess all kinds of outcomes: knowledge, skills, conceptual understanding, and "big" outcomes such as problem solving, critical thinking, group collaboration, and ability to communicate. Again, one procedure is not necessarily better than another. Each technique has advantages and disadvantages. Users need to balance the following factors:

- Purpose of the assessment—a detailed diagnosis of student ability might require more manipulatives than an overview of student ability.
- Skills and knowledge measured—science process skills might require the use of manipulatives while mathematics problem solving might not.
- Students' developmental level—younger students might require more manipulatives.
- Instructional philosophy—for example, a student-centered instructional approach might require more manipulatives.
- What the assessment communicates about what is important—if manipulatives are emphasized in the curriculum they should be emphasized in the assessment.
- Practicality—for example, it might be impossible to have every student directly design and carry out an experiment.

Users must be especially careful not to use hands-on tasks simply because students like them or they seem to be a good idea. There must be a good pedagogical reason to use them. On the other hand, we need to be aware of the message we send teachers, students, and parents by never using hands-on tasks.

### Response Mode

Different assessments require different modes of response. Most student responses are written (samples A, B, D, E, F, G, H, J, and K), some are oral (C, I), and some require production of a product (I). Note that responses can be multimodal, such as example I, in which students both produce a product and present an oral report about it.

Written responses are useful because they tend not to be bulky and they produce a permanent record. We must, however, be careful that ability to write does not mask the actual skill being assessed. Producing a product is also fairly permanent, but bulkier, and might be a better way to assess particular outcomes—for example, if you want to see whether a student can *do* something, not just talk about doing it. Oral responses might be more appropriate for young students, students with poor writing skills, and students for whom English is a second language.

### Length and Complexity

Many people talk about the "three P's"—on-demand performance assessment, projects, and portfolios. On-demand performance assessments are relatively short tasks (one to three class periods) that students do at approximately the same time and in the same manner (i.e., they tend to be standardized). Examples are D (solid waste), E (sow bugs), F (spot remover), G (aquarium), and H (bucket of beans). A type of on-demand assessment related to on-demand performance assessment is called "short answer" or "constructed response," which requires students to provide short answers, sometimes to a series of related questions (examples A, B, and C). (Some authors distinguish this type of question from their definition of performance assessment, which requires a longer, more integrated task that is more suited to assessing complicated student skills.)

Projects are longer-term activities that are embedded in classroom work and can occur more at the discretion of the teacher. An example is I (mechanized vehicle). Frequently, projects used for assessment are indistinguishable from projects used for instruction. In fact, some people maintain that projects used for assessment should model good instruction (a good discussion of this can be found in Gong, Venezky, and Mioduser 1992).

Portfolios are purposeful collections of student work that show progress over time and depth of competency. Samples in portfolios are usually chosen by students to demonstrate what they know and can do, usually in pre-established areas. For example, Sample 3 (at the end of this chapter) is a high school science portfolio for which students selected one problem-solving investigation, one work demonstrating creative expression, and one work demonstrating growth through writing; students also had to explain why they included specific pieces in the portfolio. There is often a great deal of student choice in portfolio content, and portfolios tend to have an individual look. Standardization, when it occurs, is usually at the level of the categories of items students must submit and the criteria used to judge success, rather than at a level that mandates particular work or performances for inclusion.

The options you select depend on what you want to assess and your purposes for assessment. If you need to make high-stakes, important decisions about individual students, you need to gather sufficient evidence to be sure your estimation of student ability is valid. Because some of the skills on which we judge competence might require complex behaviors, a portfolio might be called for. A portfolio allows the collection of multiple examples of complex performances over time. No single on-demand performance assessment allows this. A perfect example is the current Certificate of Mastery reform effort in several states. Under this reform, students must demonstrate mastery of complex abilities in order to receive an endorsement on their diploma. Attempts are being made to design portfolios for this purpose.

The point here is that specific types of tasks are not chosen just because they are popular or sound like a good idea. Tasks must be crafted to accomplish what is desired. If you need a quick, overall estimate of how a group of students is doing on math problem solving, then on-demand performance assessment of five problems randomly distributed among students (so that any one student gets only one problem) might be the best choice. If you need to make high-stakes decisions about individual students, a combination of approaches might be necessary.

### Group Work

Some tasks are done individually (A, B, C, D, E, F, G, and H), some are done in groups (I), and some are a combination. Group work is usually incorporated into tasks when:

1. You want to assess group collaboration. You probably will not design a separate task just to assess group collaboration; such assessment is best done in the context of other tasks.
2. You want to send a message that group collaboration is important.
3. Students need to engage in group work to "level the playing field." For example, if you wish to assess ability to write, you might assign a task for which some students have more prior knowledge than others. Group work allows for sharing background information that all students can then use to craft their essays.
4. Students need to interact in order to gather information needed for an individual task. For example, students might individually analyze survey data collected by the group.

In summary, tasks must be crafted to accomplish a vision. This vision relates to what you want to assess, the reason you are assessing it, the instructional messages you want to send to teachers, the messages you want to send to students, the relationship you see between assessment and instruction, and your instructional philosophy. This vision building is expanded in Section IV.

## Types of Performance Criteria

In alternative assessments, the quality of student performance on a task is assessed or rated using some sort of performance criteria, rubric or scoring guide. In the science and mathematics alternative assessments we've seen, the following seem to characterize the types of scoring mechanisms.

### Task-Specific Criteria vs. Generalized Criteria

Probably the major difference between scoring types is the degree to which scoring is task-specific. Task-specific scoring means that the scoring guide can be used for only one task, because you are looking for specific features in order to award points. One form of task-specific scoring is marking answers right or wrong, as in examples A, B, and C.

Task-specific scoring can also be based on assigning points for the presence or absence of specific words, calculations, phrases, or diagrams in the response. Here are examples:

**Scoring for D: Solid Waste**

*Sorting*: 7 points for "all correct." Subtract 1 point for each incorrect response. *Sequencing*: 1 point for ordering items correctly; 0 points for incorrect order. *Graph*: 2 points for labeling and ordering categories on horizontal axis (subtract 1 point for incorrect order); 7 points for correct graphic representation of each category (subtract 1 point for each incorrect representation). *Solution*: 4 points for a strong, workable solution (Lee and Van Pelt 1991).

**Scoring for F: Spot Removers**

To get a 4 the student must recognize "kinds of material," "stains," and "spot removers" as three variables to consider; describe control of material and stain, while varying only the spot remover; and (this refers to the more open-ended part of the task) clearly identify a plan that will make comparison of the spot removers possible (Kentucky Department of Education 1992).

**Scoring for Sample 1: Key Chain** (see the samples at the end of this chapter)

A "4" response contains a picture and an explanation that indicates a clear understanding of the pattern; a picture showing a whole divided into four equal parts; or an explanation that enhances the picture by comparing the size of each piece using either sentences, computations, or a combination of both.

None of these scoring guides could be used for another task. You couldn't, for example, use the scoring guide for example D to score the student responses in example F. Now consider the following *generalized* criteria:

**Scoring for Sample 2: Mathematics Problem Solving**

Performance is scored along four dimensions: conceptual understanding, procedural knowledge, problem solving and strategies, and communication. The same scoring guide is used for all problems given to students in grades 3, 5, 8, and 11.

**Scoring for I: Mechanized Vehicle**

Performance is scored for thinking skills (conceptual understanding, creativity, and critical thinking), group collaboration, and communication skills (speaking and writing). These same criteria could be used for any group project.

Other generalized criteria appear in Samples 3 (GSE science portfolio) and 4 (California Science) at the end of this chapter.

These performance criteria differ to the degree that we *indirectly* infer competence from a student's performance on a task or *directly* judge it. Although the importance of this point may not be entirely clear at first, it has far-reaching implications for how we view and use alternative assessment.

As an illustration, let's assume that we are designing an alternative assessment to measure mathematics problem solving. The *indirect inference* approach consists of developing a math problem for students to solve. Points are assigned for different features of the students' performance: getting a particular answer, having a particular graph with particular labels, using particular descriptive phrases, having a particular picture, using particular measuring instruments in particular ways—in short, task-specific scoring. Problem-solving ability is not judged directly, but inferred from the number of points the student receives. Task-specific scoring is often associated with indirect approaches of assessing skills. For instance, a high score on example G (aquarium problem) is taken as evidence of problem-solving ability.

This is essentially the same logic used in multiple-choice tests, the only differences being that in alternative assessments answers are generated by the student, not chosen from a list, and answers are given any number of points depending on particular features, instead of being scored simply right or wrong.

Contrast this approach with one at the other end of the continuum for our problem-solving illustration—direct judgment of the skill being assessed. In this approach, we design tasks to elicit problem solving as before, but then, instead of scoring the response right/wrong or assigning points for correctness or specific features, we judge problem solving directly using generalized criteria. For example, consider the four-trait model developed by the Oregon Department of Education (Sample 2). In this case, the *same* criteria are applied to *all* problems solved by students, and problem solving is judged directly by the rater.

This difference is important, because it relates to our vision of how alternative assessment fits into the educational enterprise. For example, the indirect approach (emphasizing task-specific scoring) is more reliable and faster for large-scale assessment because raters don't have to analyze each student's work for evidence of problem solving; all they have to do is look for specific diagrams, calculations, or explanations, and give points for their presence. All the judgment work has already been done. Someone else has decided that on problem X, good problem solvers would have this or that diagram, calculation, or explanation.

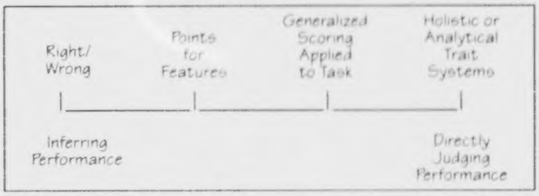
A more direct approach, emphasizing generalized scoring procedures, has important advantages for instructional use in the classroom because it *forces assessors to do the analysis themselves*. Making judgments about the qualities that make a performance successful helps assessors internalize the criteria for success. This practice becomes especially powerful if the assessors are teachers and students. The following sentiments (although addressing writing, not problem solving) explain the link to instruction:

> Scales, criteria and specific questions which students apply to their own or others' writing also have a powerful effect on enhancing quality. Through using the criteria systematically, students appear to internalize them and bring them to bear in generating new materials even when they do not have the criteria in front of them. [This procedure is] two times more effective than [other] techniques (Hillocks 1986).
>
> Winning points may be the final goal of classroom work as it is of the sports endeavor, but the grade, like the final score of the game, never taught anyone how to win again, or why they lost. For the truly successful contenders, playing the game is always about learning the game...however often it seems to be about scoring more wins than losses (Lucas 1992).

In other words, making direct judgments of the skill being assessed helps teachers and students conceptualize what "good" looks like across situations.

When scoring is done for specific features of tasks, raters get to know what "good" looks like on only a single problem; they don't have to generalize, transfer, or apply knowledge to a new situation. Therefore, direct rating of performance using generalized criteria is better as an instructional tool.

Of course, life is never black and white. What has been described is really a continuum:


Right/ Wrong
Points for Features
Generalized Scoring Applied to Task
Holistic or Analytical Trait Systems
Inferring Performance
Directly Judging Performance

In actuality, alternative assessments fall along this continuum. A happy medium can occur where a general set of scoring guidelines is developed and then systematically applied to each task to develop specific scoring guides based on points. This approach can help teachers and students systematically generalize so that eventually they aren't dependent on specific scoring guides. Such an approach is used by Kentucky (example F).

This may be a good compromise when large-scale efficiency is required or when users (including students) are learning how to apply general criteria to varied tasks. For this approach to be successful, however, the general criteria should adhere to standards of quality (comprehensive and descriptive); the link between the specific example (e.g., "has a picture showing a whole cut into four equal parts") and the general guide (e.g., "uses appropriate graphic representations") must be clear; and the ultimate goal has to be understanding the general criteria and applying them to new situations. (See Section VI in this book for additional discussion of these issues.)

### Holistic vs. Analytical Trait Generalized Scoring

Within the broader category of generalized criteria, there are also two approaches—holistic and analytical trait. Holistic scoring gives one overall judgment about the quality of the whole response (see Sample 4). Analytical trait scoring entails giving separate scores for various dimensions of performance (see Sample 2).

The advantage of holistic scoring is that it is quicker. It is used in large-scale assessment when the main goal is to get a relatively quick overview of student performance. The advantage of analytical trait scoring is that it is more diagnostic. Therefore, it is used when the goal is to provide diagnostic information to teachers to plan instruction or when the goal is to use criteria as a way for students to understand and internalize the criteria for success.

### Nature of the Scoring Scale

Once you decide on the major approach to scoring (task-specific or generalized) and whether you will have one score or several, you need to decide how many points to award. Next, you must decide what students have to do to get each score. What features make a performance excellent, good, medium or poor? Believe it or not, there *are* variations in the philosophy behind how points are assigned for each trait scored. Here are some:

**Scoring Based on Quantity.** In Baker et al. 1992, students write an essay in chemistry and are scored on four traits—concepts, argumentation, use of references to the text, and misconceptions. Three of these ratings are based on "counts"—that is, the rater counts up the number of misconceptions, concepts used, or proportion of text detail used, and the score is based on the number found.

**Scoring Based on Quality.** In Sample 4 (California science) judgments are based on the quality of the responses. For example, one feature of high performance is that "the student successfully proposes an explanation which clearly shows a relationship between data and final conclusions." The corresponding feature of a weak performance is that "the final conclusion is unrelated to the tasks performed." It doesn't matter how many conclusions the student comes to as long as the conclusions drawn are good ones.

In general, rating scales based on quantity should be avoided. One good conclusion is worth more than 10 poor ones.

### Amount of Descriptive Detail

**Scoring with Sketchy Descriptions of Terms.** In example D (solid waste), to get a high score for one portion of the exercise, the student must have given a "strong, workable solution"; however, there is no description of a strong, workable solution. What does a strong, workable solution look like?

**Scoring with Some Descriptions of Terms.** In Sample 2 (Oregon mathematics), to get a high score on procedural knowledge, the student "uses principles efficiently while justifying the solution, uses appropriate mathematical terms and strategies, solves and verifies the problem, and uses mathematical principles and language precisely." Although still not definitive, this explanation provides additional assistance with what to look for in the response to help make a judgment.

**Scoring with More Complete Descriptions of Terms.** In Sample 3 (GSE Science Portfolio), consider the additional detail provided in order to help the assessor rate the student performances he or she will encounter. For example, in the "Growth Through Writing" rubric, a strong score is indicated when "student work and self-reflection show extensive knowledge of the scientific ideas presented. The student uses relevant terms appropriately, providing evidence of excellent growth in understanding the scientific ideas through comparison of original and revised samples of work . . . ," and so on.

There are two issues related to amount of detail. First, it is easier for raters to be consistent when description is plentiful. Good descriptors are especially important for large-scale assessment. An even more important consideration than just getting good consistency between raters is the communication potential of descriptive criteria. If we tell students, for example, that their solutions have to be "workable," do we help them know what to do? Wouldn't it be more helpful to describe what "workable" means? For example: "Workable means striking a balance between being practical (in terms of time, knowledge, and money) and being effective (in terms of really solving the problem); that effectiveness has to be demonstrated by more than just opinion (you need data on the effectiveness of other similar solutions); and that the technology needs to be currently available to implement the solution." Can you think of additional details that would be useful for students?

The point here is that high-quality performance criteria can be more than just a tool for giving a score on a test. They can help us define our expectations for students and develop a shared vision of our goals for them.

In summary, many types of performance criteria are being developed. The type of scoring mechanism you use should directly relate to what you want to assess, your purpose for assessment, and how you see alternative assessment fitting into the educational enterprise.[2] Holistic and task-specific scoring are often used for large-scale assessment because they are faster. Analytical trait performance criteria having lots of description may be the best choice if you want to develop a common vision of the goals for students or if you want to increase the chance that alternative assessments will have useful instructional spin-offs in the classroom.

## Conclusion

There is considerable activity in the development of alternative assessment in science and mathematics at all levels of education (classrooms to federal governments) and at all grade levels. A variety of approaches is evident in the samples we have collected. In all these examples, the same reasons are given for changes in assessment: the need to better reflect current theories of instruction and instructional goals for students. Unfortunately, the link between the general rhetoric and the specificities of individual assessment designs is not often clear. For example, if everyone is saying we need to better model good instructional practice, why do some assessments use only manipulatives while others use only traditional word problems? Or why, when people are assessing math problem solving, do some groups use task-specific scoring while others develop general criteria? We need to step back and reconsider all our design features to make sure they really do match what we are trying to accomplish.

### REFERENCES

Baker, E. et al. (1992). *CRESST Performance Assessment Models*. Los Angeles, Calif.: Center for the Study of Evaluation; UCLA Graduate School of Education.

Bennet, D. (1992). *Assessment and Technology* (videotape). New York: Center for Technology in Education, Bank St. College of Education.

California Department of Education. (1993). *Science—New Directions in Assessment*. Sacramento, Calif.: California Department of Education.

California Department of Education. (1994). *Golden State Examination Science Portfolio—A Guide for Teachers*. Sacramento: California Department of Education.

Gong, B., R. Venezky, and D. Mioduser. (1992). Instructional Assessments: Lever for Systemic Change in Science Education Classrooms. *Journal of Science Education and Technology 1*, 157-176.

Hall, G. (1992). *Performance Assessment in Mathematics*. Alberta, Canada: Alberta Education.

Hillocks, G. (1986). *Research on Written Composition*. Urbana, Ill.: National Council of Teachers of English.

Kentucky Department of Education. (1992). *KIRIS Open-Response Items*. Frankfurt: Kentucky Department of Education.

Lee, E., and J. Van Pelt. (1991). *Discovering the Problem of Solid Waste*. Grayslake, Ill.: Lake County Educational Service Center.

Lucas, C. (1992). " Introduction: Writing portfolios—Changes and Challenges." In K. Yancey (ed.), *Portfolios in the Writing Classroom*. Reston, Va.: National Council of Teachers of English.

New Standards Project. (n.d.). *Aquarium Problem*. Pittsburgh, Pa.: Learning Research and Development Center, University of Pittsburgh.

Oregon Department of Education. (1994). *Mathematics Problem Solving Scoring Guide*. Salem, Oreg.: Office of Assessment.

Pine, J., G. Baxter, and R. Shavelson, (1991). *Assessments for Hands-on Elementary Science Curricula*. Santa Barbara,

---

[2] It is interesting that while almost all large-scale assessments have attended to developing reliable scoring mechanisms, many teacher-developed assessments tend to be sketchy in this area. Indeed, many teacher-developed assessments have no performance criteria at all.

Calif.: Graduate School of Education, University of California at Santa Barbara.

Riverside Publishing Company. (1991). *Customized Test Development Program, Performance Test Exercises*. Chicago, Ill.: Riverside Publishing Company.

Ruiz-Primo, M. A., G. P. Baxter, and R.J. Shavelson, (1993). "On the Stability of Performance Assessments." *Journal of Educational Measurement* 1: 41-53.

Shavelson, R. J., G.P. Baxter, and J. Pine, (May 1992). "Performance Assessments: Political Rhetoric and Measurement Reality." *Educational Researcher* 21: 22-27.

Whetton, C. (1991). *Science for Seven-Year-Olds in England and Wales*. England: National Foundations for Educational Research.

### SAMPLE 1

## EXAMPLE OF TASK-SPECIFIC PERFORMANCE CRITERIA

**Task:**

Mr. Ramirez helped four students make key holders. He gave Rhonda, Sam, Tony, and Uta a board and told them to share it equally. First, Rhonda measured and cut one-fourth of the board. Next, Sam measured and cut one-third of the remaining board. Finally, Tony measured and cut one-half of the remaining board. Uta used the piece that was left. Did the four students share the board equally? Draw a picture and explain your answer.

**Performance Criteria:**

A "4" response contains both a picture and an explanation indicating a clear understanding of the pattern; contains a picture showing a whole divided into four equal parts; contains an explanation that enhances the picture by comparing the size of each piece using either sentences, computations, or a combination of both.

A "3" response contains a picture that indicates a clear understanding of the pattern but only an attempt at an explanation; contains an explanation that indicates a clear understanding of the pattern but only an attempt at a picture; has limited detail in the picture or the explanation.

A "2" response offers an adequate picture only; offers an adequate explanation only; contains an explanation that does not enhance the picture; may be difficult to understand due to errors in language and grammar.

A "1" response makes some attempt at a picture or an explanation; is unclear.

Source: The Riverside Publishing Company, 8420 Bryn Mawr Avenue, Chicago, IL 60631. Reprinted with permission.

SAMPLE 2

## EXAMPLE OF ANALYTICAL TRAIT SCORING

### MATHEMATICS PROBLEM SOLVING SCORING GUIDE

**Dimensions of Mathematics**

### CONCEPTUAL UNDERSTANDING: Scoring Guide

Conceptual Understanding includes the ability to interpret the problem and select appropriate information to apply a strategy for solution. Evidence is communicated through making connections between the problem situation, relevant information, appropriate mathematical concepts and logical/reasonable responses.

5 FULL CONCEPTUAL UNDERSTANDING: The student uses all relevant information to solve the problem.

- The student's answer is consistent with the question/problem.
- The student is able to translate the problem into appropriate mathematical concepts.

3 PARTIAL CONCEPT UNDERSTANDING: The student extracts the "essence" of the mathematics in the problem, but is unable to use this information to solve the problem.

- The student is only partially able to make connections between/among the concepts.
- The student's solution is not fully related to the question.
- The student understands one portion of the task, but not the complete task.

1 LACK OF CONCEPTUAL UNDERSTANDING: The student's solution is inconsistent or unrelated to the question.

- The student translates the problem into inappropriate mathematical concepts.
- The student uses incorrect procedures without understanding the concepts related to the task.

### PROCEDURAL KNOWLEDGE: Scoring Guide

Procedural Knowledge deals with the student's ability to demonstrate appropriate use of concepts. Evidence includes the verifying and justifying of a procedure using concrete models, or the modifying of procedures to deal with factors inherent in the problem.

5 FULL USE OF APPROPRIATE PROCEDURES: The student uses principles efficiently while justifying the solution.

- The student uses appropriate mathematical terms and strategies.
- The student solves and verifies the problem.
- The student uses mathematical principles and language precisely.

3 PARTIAL USE OF APPROPRIATE PROCEDURES; The student is no precise in using mathematical terms, principles, or pocedures.

- The student is unable to carry out a procedure completely.
- The process the student uses to verify the solution is incorrect.

1 LACKS USE OF MATHEMATICAL PROCEDURES: The student uses unsuitable methods or simple manipulation of data in his/her attempted solution.

- The student fails to eliminate unsuitable methods or solutions.
- The student misuses principles or translates the problem into inappropriate procedures.
- The student fails to verify the solution.

Source: Oregon Department of Education, 255 Capitol N.E., Salem, OR 97310-0203. Reprinted with permission.

BEST COPY AVAILABLE

## MATHEMATICS PROBLEM SOLVING SCORING GUIDE—*(continued)*

### Dimensions of Mathematics

### PROBLEM SOLVING STRATEGIES: Scoring Guide

Problem Solving requires the use of many skills, often in certain combinations, before the problem is solved. Students demonstrate problem solving strategies with clearly focused, good reasoning that leads to a successful resolution of a problem.

5 EVIDENCE OF THOROUGH/INSIGHTFUL USE OF SKILLS/STRATEGIES: The skills and strategies show some evidence of insightful thinking to explore the problem.

- The student's work is clear and focused.
- The skills/strategies are appropriate and demonstrate some insightful thinking.
- The student gives possible extensions or generalizations to the solution or the problem.

3 EVIDENCE OF ROUTINE OR PARTIAL USE OF SKILLS/STRATEGIES: The skills and strategies have some focus but clarity is limited.

- The student applies a strategy which is only partially useful.
- The student's strategy is not fully executed.
- The student starts the problem appropriately, but changes to an incorrect focus.
- The student recognizes the pattern or relationship, but expands it incorrectly.

1 LIMITED EVIDENCE OF SKILLS/STRATEGIES: The skills and strategies lack a central focus and the details are sketchy or not present.

- The procedures are not recorded (i.e., only the solution is present).
- Strategies are random.
- The student does not fully explore the problem looking for concepts, patterns or relationships.
- The student fails to see alternative solutions that the problem requires.

### COMMUNICATION: Scoring Guide

In assessing the student's ability to communicate mathematically, particular attention should be paid to both the meanings he/she attaches to the concepts and procedures and also his/her fluency in explaining, understanding, and evaluating the ideas expressed.

5 CLEAR, COMPLETE COMMUNICATION: The student gives a complete response with clear, coherent, unambiguous, and elegant explanations.

- The student communicates his/her thinking effectively to the audience.
- The details fit and make sense.
- One step flows to the next and shows organization.
- The student presents strong supporting arguments.

3 PARTIAL OR INCOMPLETE COMMUNICATION: The student's explanation is unclear, inconsistent or not complete.

- The student uses terminology incorrectly or inconsistently.
- The student's visual aids (graphs, tables, diagrams, etc.) are inappropriate or not directly related.
- The student's explanation centers on his/her solution, not on his/her thinking.

1 LIMITED OR LACK OF COMMUNICATION: The student's explanation is not understandable or not present.

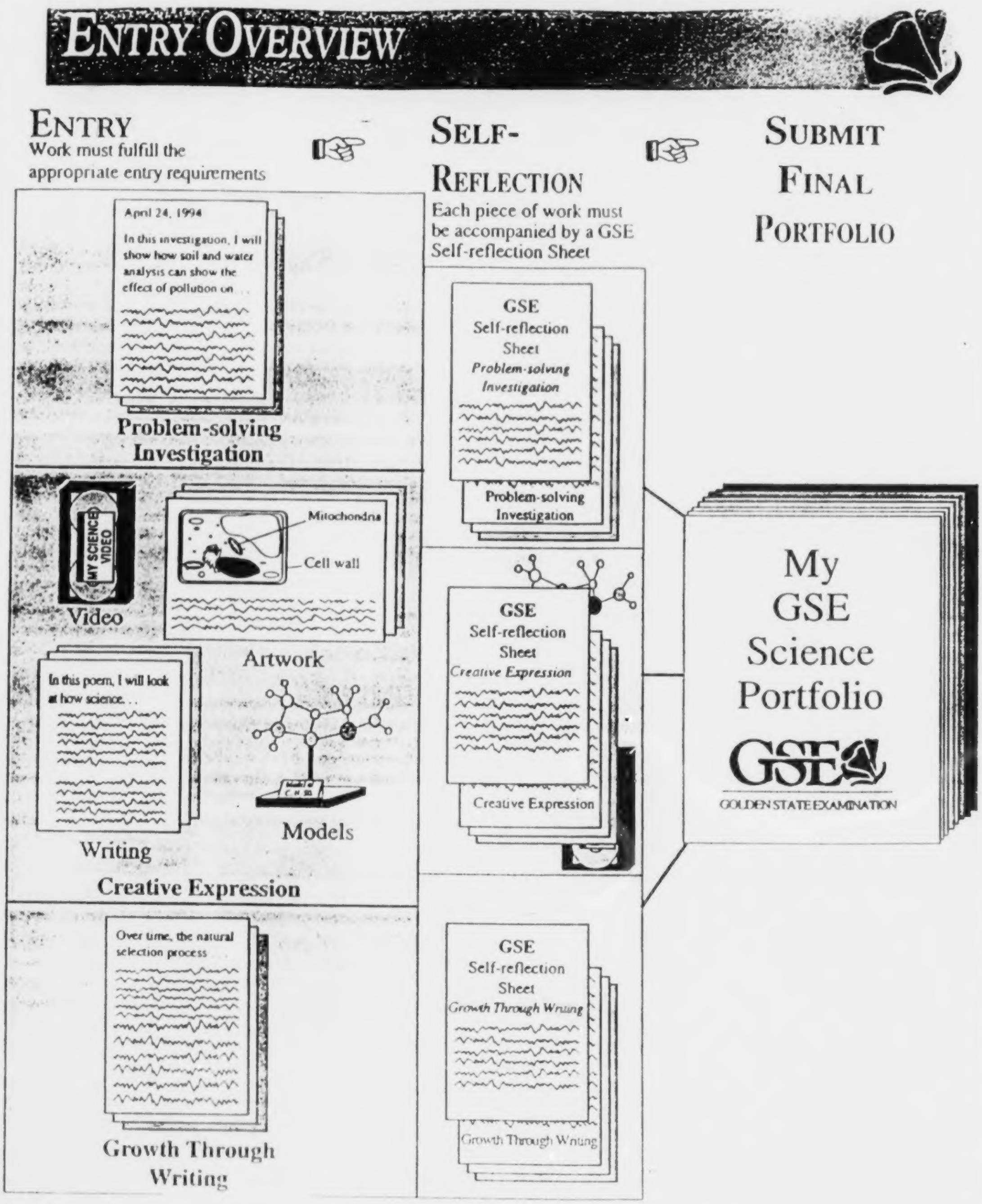
- The student either does not use or misuses appropriate mathematical terminology.
- The student does not use essential visual aids to enhance or clarify the explanation.
- The student's explanation lacks focus.

*Source:* Oregon Department of Education, 255 Capitol N.E., Salem, OR 97310-0203. Reprinted with permission.

BEST COPY AVAILABLE

SAMPLE 3

GOLDEN STATE EXAMINATION SCIENCE PORTFOLIO



ENTRY OVERVIEW

ENTRY
Work must fulfill the appropriate entry requirements

SELF-REFLECTION
Each piece of work must be accompanied by a GSE Self-reflection Sheet

SUBMIT FINAL PORTFOLIO

April 24, 1994

In this investigation, I will show how soil and water analysis can show the effect of pollution on . . .

In this poem, I will look at how science . . .

Over time, the natural selection process . . .

Problem-solving Investigation
MY SCIENCE VIDEO
Video
Mitochondria
Cell wall
Artwork
Writing
Models
Creative Expression
Growth Through Writing
GSE
Self-reflection
Sheet
Problem-solving Investigation
Creative Expression
Growth Through Writing
My GSE Science Portfolio
GSE
GOLDEN STATE EXAMINATION

BEST COPY AVAILABLE

SAMPLE 3—*(continued)*

## GOLDEN STATE EXAMINATION SCIENCE PORTFOLIO

# ENTRY SCORING GUIDE

## *Problem-solving Investigation*

*Level 6* Student work and self-reflection show the student's exemplary skill in experimental design. Excellent observations and data analysis indicate extensive knowledge of the scientific ideas presented. Analyses and conclusions are supported by observations and data and show a high level of reasoning. The entry includes conclusive evidence that the student worked cooperatively with others. The self-reflection sheet clearly identifies how working with others improved the student's understanding of the scientific ideas presented. The student's investigation makes insightful applications to real-world situations. All aspects of the task and analysis are completed thoroughly. Written expression is exemplary and contributes to clear, coherent and communication.

*Level 5* Student work and self-reflection show the student's strong skill in experimental design. Observations and data analysis indicate very good knowledge of the scientific ideas presented. Analyses and conclusions are supported by observations and data and show a high level of reasoning. The entry includes strong evidence that the student worked cooperatively with others. The self-reflection sheet clearly identifies how working with others improved the student's understanding of the scientific ideas presented. The student makes generally valid applications to real-world situations. All aspects of the task and analysis are complete. Written expression is very good and contributes to clear and coherent communication.

*Level 4* Student work and self-reflection show the student's solid skill in experimental design. Observations and data analysis indicate solid knowledge of the scientific ideas presented. Analyses and conclusions are supported by observations and data and show a moderately high level of reasoning. The entry includes solid evidence that the student worked cooperatively with others. The self-reflection sheet identifies how working with others improved the student's understanding of the scientific ideas presented. The student's investigation makes generally valid applications to real-world situations. All major aspects of the task and analysis are complete. Written expression is good and contributes to coherent and effective communication.

*Level 3* Student work and self-reflection show the student's basic skill in experimental design. Observations and data analysis indicate basic knowledge of the scientific ideas presented. Analyses and conclusions may be supported by observations and data. The entry attempts to provide evidence that the student worked cooperatively with others. The self-reflection sheet identifies how working with others caused some improved understanding of the scientific ideas presented. The student attempts to make some applications to real-world situations. Some aspects of the task and analysis are incomplete. Written expression is flawed and shows basic communication.

*Level 2* Student work and self-reflection show the student's limited skill in experimental design. Observations and data analysis indicate limited knowledge of the scientific ideas presented. Analyses and conclusions may not be supported by observations and data or may be incomplete. The entry may attempt to provide evidence that the student worked cooperatively with others. The self-reflection sheet may identify how working with others caused limited improvement in the student's understanding of the scientific ideas presented. When made, applications to real-world situations are limited. Tasks and analysis are largely incomplete. Written expression is flawed and shows limited communication.

*Level 1* Student work and self-reflection show the student's minimal skill in experimental design. Observations and data analysis, if present, indicate minimal knowledge of the scientific ideas presented. Analyses may be incomplete and conclusions may not be supported by observations and data. The entry shows little evidence that the student worked cooperatively with others. The self-reflection sheet may identify only minimal improvement in the student's understanding of the scientific ideas presented. When made, applications to real-world situations are minimal. Tasks and analysis are mostly incomplete. Written expression is very flawed and is contributes to ineffective communication.

© 1994 Golden State Examination Science Portfolio
California Department of Education

Permission for use in training has been granted by
the California Department of Education. Materials may be
copied for training purposes only.

### SAMPLE 4

## EXAMPLE OF HOLISTIC SCORING GUIDE FOR SCIENCE

Draft Scoring Rubric—Grade 10 Performance Assessments

The following rubric was developed to assist teachers in scoring student responses to performance-based assessment items. This rubric allows for a wide range of student responses and should be used as a general guideline. No one criteria [sic] should be given more weight than another.

**Level 4**

The student demonstrates a clear understanding of the problem and investigations performed. The student's observations are valid and demonstrate attention to detail. The quality of the data collected reflects their success at performing each experiment. The design of experiments suggested by the student are well thought out and scientifically accurate. All data is [sic] presented in an organized fashion, utilizing appropriately designed tables and charts. Answers to questions are well thought out and supported by experimental data. In general, there are no false assumptions or misleading statements made by the student. The student recognizes the need for additional testing and provides appropriate suggestions related to the problem. The conclusions reflect the student's ability to effectively analyze experimental data and draw appropriate conclusions. The student successfully proposes an explanation which clearly shows a relationship between their [sic] data and their [sic] final conclusion.

**Level 3**

The student demonstrates an understanding of the problem and investigations performed. The student's observations are valid and demonstrate some attention to detail. The quality of the data collected generally reflects their success at performing each task. The experiments suggested by the student are well thought out, but many have flaws in scientific design. Most data is [sic] presented in a organized fashion, utilizing appropriately designed tables and charts. Answers to most questions are well thought out and supported by experimental data. In general there are few false assumptions or misleading statements. The student recognizes the need for additional tests, but is unable to provide appropriate suggestions related to the problem. The conclusions reflect the student's ability to analyze experimental data and draw conclusions. The student successfully proposes an explanation which generally shows a relationship between their data and their final conclusion.

**Level 2**

The student demonstrates some understanding of the problem and investigations performed. The student's observations are vague and lack attention to detail. The quality of the data collected reflects limited success at performing the tasks. Some tasks may have been omitted. The experiments suggested by the student are unclear or riddled with flaws in scientific design. The data is [sic] recorded in a disorganized fashion, with tables and charts that are poorly designed or missing. Answers to questions are not well thought out or supported by experimental data. In general, the student makes both false assumptions and misleading statements. The student fails to recognize the need for additional tests. The student has difficulty analyzing experimental data and drawing conclusions. The student fails to propose an explanation which shows a relationship between their data and their final conclusion.

**Level 1**

The student demonstrates little or no understanding of the problem or investigations performed. The student's observations are poor or missing and show no attention to detail. The quality of the data collected reflects little or no success at performing the tasks. The experiments suggested by the student were difficult to follow, or missing. The data collected is [sic] haphazardly recorded, or missing. Answers to questions are implausible and not related to the experimental data. In general, the student makes many false assumptions and misleading statements. The student fails to recognize the need for additional tests. The student in unable to analyze experimental data or draw conclusions. The final conclusion is unrelated to the tasks performed, or missing.

*Source: Science—New Directions in Assessment*, 1993, California Department of Education, 721 Capital Mall, Sacramento, CA 94244.

BEST COPY AVAILABLE

# V-4

# Keys to Good Writing Prompts

Vicki Spandel
Northwest Regional Educational Laboratory

Let's be very honest here: the perfect prompt doesn't exist. No writing topic, no matter how clear, how stimulating, how cleverly written, how structured or open, how inviting, will guarantee for every student the maximum chance of writing success. Acknowledging that bit of collective wisdom, we can assert that some prompts are ever so much better than others—better in the sense that they welcome students to the task of writing, allow them the freedom to show what they can do as writers, and encourage the sort of writing that holds raters' attention, thereby enabling them to focus on the quality of the writing rather than on an annoying or misleading topic.

Even if 20 years of intensive work in the field of direct writing assessment have not yielded the prompt that works for all writers in all situations, the experience has given us time to identify specific guidelines for developing *workable* prompts. This paper outlines, briefly, several keys for developing good prompts and also reviews some general strategies for handling the multifarious—and often unpredictable—ways writers interpret prompts.

## Keys to Creating Workable Prompts

### Key 1: Let the Writer Draw on Experience

Every experienced writing teacher knows that student writers are at their best when they can and do use what they know, yet many prompts preclude this. A prompt asking students to identify major causes of pollution within a community, for instance, demands specialized knowledge that simply makes the task too hard. The people who write the prompts are not always the same ones who administer or score the writing tests. It may not always be apparent to prompt developers that in a typical writing assessment, there is no time or opportunity to visit the library or even glance at an encyclopedia, much less do thorough research.

What is needed for large-scale assessment purposes (even with an expository or persuasive writing task) is a topic students already know something about, even if they cannot back their assertions with hard data. Here's an example of an expository prompt that satisfies those criteria: *"What qualities do you think a hero should possess? Write a paper that EXPLAINS what it means to be a hero. You may wish to use someone from literature or public life or someone you know personally as an example."* Notice that this prompt invites students to use what they think and believe, as well as what they've gained from reading or interacting with others.

Allowing writers to use what they know also demands sensitivity to experiential bias. Not all student writers live with two natural parents, have brothers and sisters, travel, attend films, keep pets, watch television, or visit large cities (or the country); yet many prompts require just this sort of experience in order for the writer to respond. In writing prompts, we must think of what is likely to be within the usual, everyday experiences of most student writers, given their age, culture, geographic location, range of experience, and school environment. Even then, assuming universal experience is never safe; this is a good argument for offering a choice of prompts.

### Key 2: Allow Choice

Choice opens mental doors for students and there are several ways to achieve it. One is with a prompt in which the structure and content are rather open, allowing the writer to go in whatever direction feels right: *"Think of a place you have visited that is so vivid in your mind you can almost feel what it is like to be there. It might be as large as a city or as tiny as a corner of one room. DESCRIBE this place so clearly that the reader can experience it with you."*

One student, responding to this prompt, wrote, *"The sweat on your brow. A layer of dust on your face. Out in the woods. Somewhere. On a horse. Of all the places in the world, I feel best on a horse . . . . . . ."* Notice the liberal interpretation of the word *place*; this student isn't writing about the beach or Disneyland or any of a dozen cliched responses that spring to mind. She has found her own path. Other students responding to this same prompt envisioned such diversely intriguing places as the tree in the back yard, grandma's rocking chair, a sauna, a pitcher's mound, and *"within my own mind."* Notice that none of these responses is very conventional. But isn't this what we want students to do, to think independently and write something individual? This prompt is, in fact, worded in a way that invites interpretation.

Another way to create mental elbow room is to allow a choice of topics, either within one prompt or by giving the student two prompts to pick from. For instance, the prompt *"Think of a time you felt frightened; tell what happened"* is very restrictive in comparison to:

*"Think of a time you felt frightened,*
*frustrated,*
*annoyed,*
*nervous,*
*embarrassed,*
*satisfied, OR*
*surprised.*

*Pick ONE, and tell the story of what happened."*

A third alternative, of course, is to offer the student two very different yet comparable prompts to choose from. For instance, a student might be asked to describe a *memorable place* OR a *memorable person*. These tasks are comparable in difficulty, yet one may be much more stimulating to a given writer than the other. The writer may also be given the option of coming up with his or her own topic; this is an option teachers of writing often argue for, and it sounds appealing in theory. In fact, however, self-selecting a topic often works better in classroom writing than in large-scale assessment. Within the familiar walls of the classroom or writer's workshop, the student is likely to be more relaxed, have more time for prewriting and brainstorming ideas and possible topics, and have more opportunities for discussing possible ideas with others (teacher, peers, family, friends) than exist in most formal writing assessments. Further, students accustomed to coming up with their own topics may find this a natural and desirable way to approach writing; students who are used to having writing assigned, however, may actually be stymied when faced with the prospect of charting their own writing destiny. Prompt development is tough—even if you're doing it only for yourself.

### Key 3: Offer a Prompt That Will Spark the Writer's Imagination

Easy to say. Hard to do. Obviously, no one prompt will interest *everyone* (another reason for allowing choice), but some prompts interest almost no one. Here's an example: *"Write a paper describing your favorite meal."* A teacher-student survey distributed with this prompt indicated that most of the students who responded found *favorite meal* an incredibly dull topic. And their writing—voiceless, bland, virtually detail-free—confirmed this lack of involvement.

A good way to tell if a prompt will work is to try it out; prompt developers should do this routinely, even if it sounds time consuming and tedious. After all, if the people who thought up the topic don't want to write on it (or find it difficult to begin), students can hardly be expected to pounce on it enthusiastically. Prompt developers apparently didn't try out the favorite meal prompt or they would quickly have put it through the shredder. Perhaps they were hoping for original recipes and adventures in exotic cuisines; but they didn't get it—not from the 5th and 7th graders who struggled with this bleak topic. Instead, they got hundreds of look-alikes endlessly and inevitably recounting the virtues of juicy steak, crispy pizza, fluffy potatoes, and crunchy chicken. It wasn't what they wanted—but it *was* what the prompt called for.

By contrast, here's a prompt that did stimulate imaginative and highly diverse responses: *"[Students are presented with an old shoe—or a choice of several shoes.] Think who this shoe might belong to and how it came to be found alongside the road in your neighborhood. Write the story of what happened."* Virtually all responses to this prompt differed from one another, making the readers' task not only infinitely more enjoyable, but much simpler as well. The notion that it is easier to assess writing when responses look and sound alike is a myth; in fact, it becomes almost impossible, after a time, to maintain the perspective that tells a rater what is exceptional idea development and what is not. If we want diversity, we must write prompts that not only permit but actually encourage such diversity.

### Key 4: Don't Do the Thinking for the Writer

Often, in a well-intentioned effort to be supportive, prompt developers go too far with suggestions, burying writers in a deluge of tips for success. Some writers manage to sort through the barrage of helpful hints, but too many, struggling to please, try to respond to everything, and wind up with a paper that reads like a frazzled collection of short answers. Remember that *vivid place* prompt—*" . . . big as a city, tiny as the corner of a room"* ? Here's another version of the same prompt, bogged down by over-written directions:

> *"Think of a place that is special to you, a place you have been that you might like to return to. It might be a place indoors or out, a large place or a small one, somewhere near your home or far away. Think of the sights, sounds, and smells that make this place so memorable. DESCRIBE it in a way that your reader will find it memorable, too."*

Imagine student writers trying desperately to fulfill all the expectations embedded in these complex directions, and you will have a pretty good notion of the kinds of papers such a prompt usually produces.

A good rule of thumb is this: don't make any suggestions you do not want the student writer to follow *quite literally*. Suggestions are frequently interpreted as mandates. That's why responses to complex directions often read more like fill-in-the-blank summaries than cohesive, coherent, focused pieces of writing.

Akin to thinking for the writer is the tendency to invest the prompt with creativity or cleverness so that the writer won't have to be creative or clever herself. This rarely works. In fact, it often backfires by actually blocking the writer's thinking.

Here's one example of a prompt intended to provoke creative thinking—but that in reality stifled imagination: *"Imagine that you woke up one day and you had turned into jello. Write the story of how you would get through the day."* This prompt resulted in some blank papers and a few notes like this one: *"I just cannot think of anything to write about this. I'm sorry, but I think it's stupid."* For some students, the idea of turning into jello might sound like great fun, but others simply do not think this way, and would not play a "let's pretend" kind of game to show off their imaginative skill. In all fairness, of course, the prompt writers weren't trying to annoy their student victims; they simply assumed—as prompt writers often do—that a prompt which had enjoyed some success in a classroom (with lots of teacher preparation and support and the right atmosphere) would also work in a district assessment. Unfortunately, this is often not the case. Here, then, is a corollary to Key 4: When moving a prompt from the classroom to the large-scale assessment level, be certain that it is understandable to student writers without significant explanation, that it is a prompt with widespread appeal (not everyone will know the teacher from whose class it came), and that it is not dependent on teacher-led prewriting to work.

### Key 5: Avoid Built-in Positives and Negatives

Many prompts begin with language like this: *"Think of a favorite place . . . ," "Think of a friend you liked that you will always remember . . . ," "Think of a day you enjoyed a wonderful adventure . . . ."* But *favorites, bests,* and *most wonderfuls* do not always make for the most interesting writing. Such language often prompts a student writer to adopt a "nice guy" persona that can result in stilted, artificially congenial prose.

By contrast, consider the student who wrote, *"It was the year 1985. I was in the third grade, and I had the psycho teacher from hell . . . ."* That opening line has hooked many a reader. What follows is a riveting and memorable description of an unforgettable character who made quite an impression on her 3rd grade "hostages." This is not a pleaser paper, fearful of offending. It is every inch the student's own, and it resounds with voice. It was written in response to the prompt *"Think of a teacher who would be hard to forget. . . ."*

Phrases like *memorable* or *hard to forget* give students license to be honest and open in their responses to life, rather than channeling their feelings and thoughts in ways they believe prompt writers and teacher-raters will deem appropriate. The best writing is honest and seeks its own labels and descriptors. Given the freedom to go in any direction, some students will write of treasured friends, favorite teachers, and places they love. Others will take a different approach, but all will come closer to crafting something truthful and therefore eminently readable.

### Key 6: Don't Be Nosy

In an effort to keep prompts from being boring, we may sometimes make them too controversial—or word them in ways that intrude upon students' privacy. This practice can be risky for several reasons. First, and most obviously, students and parents alike may resent prompts that ask personal questions about individual beliefs and actions or about family life, for instance: *"Almost everyone would like his or her parents to do something differently. How would you like to change your parents if you could?"* This incredibly prying prompt, used in several districts a decade or so ago, probably wouldn't survive the review of most bias or assessment ethics committees today. Although we may shrink from its intrusiveness, this prompt is rescuable, with some attention to wording. Suppose it were rewritten to read this way: *"Many qualities are important in being a good parent. Think of one, two, or three qualities you think are essential to good parenting and EXPLAIN what it means to be a good parent. Use examples from films, books, or real life to help make your ideas clear."* Now the writer is free to share personal experience or not.

Intrusiveness isn't the only problem with poorly worded prompts. Some prompts seem to come equipped with built-in expectations, like this one: *"Many people have noticed that the drop-out rate is increasing, especially in recent years. Take a stand for or against staying in school, and give reasons that would convince someone to agree with you."*

The prompt says *for or against*, but most students will assume that responses which argue for staying in school will meet with more approval than those which take a less popular view. They may or may not be right. The point is that assumption drives students' writing, often resulting in a predictable litany of platitudes that may or may not reflect their real thinking.

We can avoid some predictability with a more open-ended question like this one: *"A grant of $25,000 has been made available to your school. The School Board can decide to spend the money (1) buying books for the library, (2) funding a field trip to Washington, D.C. for senior students, or (3) for another project if a creative and educationally sound idea is suggested and the Board feels it deserves funding. Decide how you think the money could best be spent and write a paper convincing the Board to go along with your idea."* The less predictable the desired response, the more original and honest the writing is likely to be.

Finally, of course, controversial topics may result in rater bias. Teacher-raters, like everyone else, have favorite topics and often feel passionate about current social or political issues. It is hard for a rater to fairly evaluate development, organization, and voice in a paper promoting ideas that violently conflict with her personal philosophy. Why put raters to the test? Controversy is part of education, certainly, but it is better dealt with in contexts where a conflict of ideals cannot skew grades or scores.

### Key 7: Make the Purpose Clear

In many assessments, the accepted philosophy seems to be that writing is writing, and no one much cares whether students craft stories, personal essays, formal essays, persuasive arguments, or descriptions. Fine. But if purpose counts, then it is important to make that clear.

For instance, if the expectation is that students will write stories, then the words *tell the story* or *tell what happened* will usually lead the writer into this form. Similarly, the word *explain* usually triggers an expository response, while *persuade* or *convince someone to agree with you* . . . tends to elicit a persuasive argument. This doesn't always happen, of course. Given a choice, and without a reason to believe that form matters (or that it will affect their scores), most students will write personal narrative. Some may not understand the difference between using a narrative anecdote to *enrich* an essay or argument and slipping permanently into narrative mode. What is to be done about this?

First of all, we must come to grips with whether this is an issue at all. Do we care, for the purposes of large-scale assessment, whether students write in various modes or forms—expository, narrative, descriptive, persuasive? If the answer is *no*, then we should purposefully write prompts that allow students freedom to adopt the form they believe is most suited to what they're trying to accomplish as writers. If the answer is *yes* (and it may not be for everyone), then two things probably need to happen.

First, we must accept the fact that making students aware of the underlying differences among various forms of writing goes well beyond the bounds of what can normally be accomplished during formal assessment. To mentally distinguish among modes, students need an in-depth understanding of different purposes for writing, and some training in reading prompts to ascertain intent—that is, to figure out what the prompt writer probably was looking for. Such in-depth instruction needs to occur at the classroom level; a quick definition of expository writing, provided just at the moment a student is expected to begin drafting an expository essay, will not suffice. Any expectation that it will is unrealistic.

Second, even assuming the best classroom instruction on modes and purposes of writing, writers of assessment prompts and directions have explicit responsibility for making sure that all expectations regarding purpose are clearly defined. As noted previously, prompts can open with cue words that tend to naturally guide alert writers into certain modes—even when they have not had much formal instruction in modes of writing. In addition, prompt writers can give explicit directions or criteria:

*Do you think the driving age in our community should be lowered from 16 to 15? Write a paper that EXPLAINS the probable impact of such a decision on various groups within the community. Use a story as an example, if you wish, but be sure that your paper EXPLAINS the impact on our community of lowering the driving age.*

*Remember, that a good EXPOSITORY (or EXPLAINING) paper*

- *Has a clear main idea or ideas*
- *Provides clear support for each main idea*
- *Goes well beyond the obvious or what everyone already knows*
- *Makes the reader think*
- *Uses stories only to illustrate key points*
- *Gives the reader important information he/she needs and wants.*

Good, clear criteria like these support and round out the expectations defined within the prompt itself—just one more way of increasing students' chances for success.

## Some Final Thoughts and Strategies

A good prompt invites the writer—and the reader—to think, allows freedom of imagination, guides just enough but not too much, and makes readers eager to see how students have responded. What a lot to ask. Wouldn't it be nice if we could just use a formula and come up with a good prompt? It's not out of the question; in fact, some prompt developers have tried just that. Here's one possibility: an opening statement; a directive to write, based on that statement, specifying the purpose for writing; and a clarifying sentence with one or two general hints that offer helpful guidance without locking the student in. A prompt developed using this formula might look like this:

(Opening) *Think of an event you will want to remember when you are old.* (Directive specifying purpose) *TELL THE STORY of what happened in a way that is so clear that someone reading your story could tell it back just as if it were their own.* (Helpful hint) *Try to include details that surprise the reader.*

This is a good prompt with possibilities to engage students and bring out the best they can do. But this is a good prompt, too:

*Make up a story using ONE of these words:*
*Survival*
*Frustration*
*Challenge*

It doesn't follow the formula, but it erects no barriers to free expression, it allows and even encourages diversity, and it nudges the imagination. These are the things any good prompt must do. The point is, we don't need formulas to ensure good prompts. They may be helpful, but what really matters is sensitive attention to the keys of sound prompt development.

The way a rater interprets a response is as important as the student's response to a prompt. This interpretation goes beyond the actual task of evaluating. Through skillful training, raters can be taught to assess such qualities as organization, word choice, or use of conventions with impressive consistency and accuracy. Such consistency can be quickly undermined if raters approach responses from philosophically different positions. Consider the student who, responding to the favorite place topic, wrote about how it felt to be on top of a horse. For some raters, such a response signals imagination, originality, individuality—all of which are likely to be rewarded with high scores. But others may feel such a response is a little off track, that the back of a horse is not, strictly speaking, a place in the same sense that, say, the San Diego Zoo or the Olympic Rain Forest would be considered a place. And consider that this is only a *little* deviation from the ordinary. Another student, thinking of grandmother's farm as a favorite place, may be sidetracked into writing a narrative about the time she first drove her grandfather's pickup truck. Does this matter? Is it a good thing, even a desirable thing, for student writers to follow wherever the thought process takes them? This is the question, for if the rater feels that the author of the pickup truck paper failed to follow directions, her scores may suffer accordingly.

Nor is it the scores alone that matter. This fundamental difference in interpretation underlies the whole way we approach writing assessment. At issue is the question of whether we wish to assess students' ability to follow directions precisely (which many raters and teachers consider significant and many consider trivial), or whether we wish to encourage writers to shape a topic to fit their own experience and world view and interests. The latter philosophy tends to view a prompt as just that—a stimulus to write, nothing more. This interpretation gives the writer lots of room to explore and to leap from one idea to another. Of course, if we adopt the view that students should think expansively and view writing as a mental journey with no predetermined destination, we must also provide instruction which gives them experience in doing just that.

A final thought: Good prompts are important—even critical—but it is a mistake to imagine that a student's chances for success in writing assessment ride solely on the prompt. Like it or not, at some point we must stop blaming the prompt for all that goes wrong in assessment. The truth is that skillful writers *do* learn to rise above the prompt. They find ways to personalize a writing task, whether asked about a favorite object, memorable teacher, or best meal, and their writing shines with individuality. Part of the challenge to us, as teachers of writing, is to help students find ways of coping with less-than-perfect prompts, and making their responses more complete, thoughtful and inventive than writers of favorite meal and jello prompts have a right to expect. Combine that skill with thoughtful, careful prompt writing, and we will see better writing from our students than we ever thought possible.

# V-5

# Using Scoring Considerations to Develop Effective Performance Assessment Tasks

GAIL LYNN GOLDBERG
MARYLAND STATE DEPARTMENT OF EDUCATION

As schools, school systems, and states turn to more authentic modes of assessment, efforts to develop meaningful and effective performance assessments are often characterized by linearity rather than recursiveness. This linearity is manifest in a sequence of steps: (1) identification of learning outcomes; (2) development of assessment tasks intended to demonstrate attainment of those outcomes; and (3) development and application of a scoring protocol using rubrics, checklists, or other scoring tools. Scoring, however, is not merely an event that follows task development and administration, during which data are obtained on student performance. Rather, it should be regarded as a complex process that shapes what we look for in, and can say about, student performance. Experience in scoring performance assessments offers important lessons about effective task development. By considering these lessons, one can avoid or eliminate task features that may generate problematic responses—those that do not yield valid, reliable evidence of what students understand and can do and thus defeat the purposes of the assessment.

We have found the experience of scoring student responses to the Maryland School Performance Assessment Program (MSPAP) particularly instructive for subsequent task development. The MSPAP is administered annually to all students in grades 3, 5, and 8 and features content area-specific and integrated performance tasks. From these tasks, data are obtained on school and district-level performance in reading, writing/language in use, mathematics, science, and social studies. The performance tasks are paper-and-pencil, constructed response format, centering on a specific topic, problem, or investigative context. Each MSPAP task comprises a set of activities intended to measure one or more of the state board-adopted learning outcomes in these content areas.

While sources of scoring-related problems may be infinite, through four years' experience in developing and scoring performance assessment tasks for MSPAP, we have identified some of the most common ones. By introducing task writers to these potential pitfalls early on in the development process, we have significantly reduced both the revision needed after field testing and the frequency of problematic responses. Occasional problems with task wording and format are by no means unique to MSPAP, however, and neither are the remedies for them. Attention to the scoring considerations that follow can help other educators evaluate and improve the effectiveness of activities they are formulating or planning to use for assessment purposes.

## Scoring-Related Questions and Recommendations

1. **Whenever possible, have you combined task activity directives (e.g., describe, explain, list, identify) to avoid partial responses that attend to only one of a sequence of such directives?**

Experience has shown us that students, across grade levels, tend not to read questions thoroughly. Often, they springboard off the first cue they see, and they do not address all the instructions in a given activity. Recognizing that an important skill which we desire students to attain is the ability to "unlock" questions, we may not wish to confound evidence of content knowledge with reading ability. Thus, instead of asking "In which region did the drought occur? Describe this region" (sequential directives), we ask students to "Identify and describe the region in which the drought occurred" (compound directive). Awareness of this issue permits task developers to make conscious decisions about the frequency of activities with compound versus sequential directives, in relation to content areas and grade levels being assessed.

2. **Have you used task activity directives that yield a scorable product?**

This question often catches developers of paper-and-pencil performance assessment tasks off-guard, since they assume that they are creating activities that will generate a written response and therefore can be scored. Experience has shown us, however, that students may respond to some task directives with an unmeasurable behavior. For example, when presented with the directive "Based on what you have read and what you know, pick a position and explain why you think it is the most appropriate one," students will often embark immediately upon an explanation. "Pick," "choose," "think about," and "consider" are all mental operations that do not necessarily require identification or explication. Some students do not make distinctions between everyday classroom and testing contexts and therefore respond as they would in the more familiar discourse situation. While a classroom teacher could easily ask for clarification of the intended position or choice, in the large-scale testing context there is no such opportunity. Therefore, we must be more explicit about what ideas or information we want a response to include.

Conversely, we must be sensitive to the weight that activity directives, or words perceived as such, can have for some students. This became apparent when students responded to a series of MSPAP activities on workers in a community. After designing a web to show the interdependence of community members, students were asked several steps later to "Choose one worker to remove from the web. Write that worker and the job that worker does." While the intent of task writers was to lead students ultimately to consider, hypothetically, the impact on the community of not having that job done, many students dutifully responded by erasing one of the workers whom they had identified in the previously developed workers' web. Imagine the consternation of scorers as they attempted to read the tracings left behind after erasure!

3. **Whenever development or extension of ideas or opinions is desired, have you raised questions in such a way as to avoid a legitimate response that precludes such elaboration, extension, or explanation?**

One of the steps in a pilot task issued to Maryland teachers as a model for instructional purposes asked, "Is there any other information you think you would need to build an anemometer?" Too often, task writers develop questions with the intended or "right" answer in mind, one which, in this case, would require a developed response. Whether they have a poor understanding of the scientific principles involved or have insufficient prior knowledge, some students will simply respond "no" to such a question. As a general rule of thumb, I suggest that task writers include yes/no questions only when either response is conceivable, based on the support or explanation that students are asked to provide.

4. **Have you avoided phrasing questions in such a way that they can generate a response with no evidence of the learning outcome(s) being measured?**

The potential problem that this question targets is closely related to the preceding one, since in both instances students can indeed respond in a way that is feasible, but essentially unscorable; that is, while readers may recognize on a gut level that the response is legitimate and sensible, it provides no evidence that a given learning outcome has been attained. Consider, for example, asking students who have finished reading *Little House on the Prairie*, "How did the illustrations help you to understand the story?" It is possible that a student might respond, "They didn't, because I already had a good idea what the characters looked like and did from watching the television series." Certainly that must be considered a legitimate response, yet we have no evidence of construction, extension, or examination of meaning, evidence of which are needed to assign an appropriate reading score. Changing the wording to "How might the illustrations help someone understand the story?" would avoid this pitfall.

5. **When developing an activity to which students can respond in alternative ways (e.g., by taking a positive or negative approach or expressing a favorable or unfavorable position), have you made sure that development of all alternatives is possible?**

Task writers sometimes develop what they believe is an "open-ended" activity that turns out to be clearly skewed. For example, asking students "Are the directions clear enough? Explain why or why not" may not yield the evidence necessary to obtain a reading measure if the response is affirmative. For example, a student may simply respond, "Yes, I can follow them easily." Where position is not relevant, but the synthesis of ideas and information from a text *are*, we ought ask them instead to "Explain why or why not, using information or examples from the reading text." After all, if what we are seeking is the right or best answer, we can probably develop multiple-choice items to provide data in a less costly and time-consuming way. One benefit of performance assessment is the diversity of possible responses; evidence of attaining the essential outcomes being measured—not a predetermined position or approach—is the basis for our judgments.

6. **Have you avoided inter-item cuing? That is, in a series of steps or activities, have you made sure that there are no means for students to go back and change any previous step(s) that you intend to score?**

Inter-item cuing, a pitfall in developing multiple-choice items, also surfaces periodically in the development of open-ended performance assessment activities when these are sequential components of a larger task. Particularly because of scaffolding often built in to provide context, task writers sometimes inadvertently provide substantive clues about the direction responses should take. For example, after identifying fireworks as a symbol of celebration in an earlier activity, a social studies task on symbols of freedom directly cued students by asking "When might you want to set off fireworks? Explain why." We shouldn't be surprised when many students respond, "I'd set off fireworks on the last day of school, 'cause then I'll be free!"

Sometimes, the problem of inter-item cuing occurs because of the need to obtain discrete measures of given outcomes that may be evidenced only in concert with others. So, for example, an activity designed to provide evidence of estimation ("Look at the key and estimate how many miles it is from New York to Washington") may yield an invalid measure of that learning outcome because subsequent activities within the same task (i.e., "Now measure and calculate that distance") make the estimation redundant by cuing for a feasible response. The interrelationships among component activities like these must be addressed to obtain meaningful information about student performance. Also related to this review strategy is learning to avoid wording activities in such a way that students could legitimately respond by citing a previous response, and instead cuing students to provide new, related, or more developed ideas or information.

7. **Have you used response "completeness" cues with care and clear intent? Specifically, have you either indicated a specific number of reasons, examples, or items of information or cued instead for development so that multiplicity of ideas and extension of ideas are equally acceptable?**

I often have to impress upon task writers a cardinal rule of developing scorable activities: we should not evaluate that which we haven't told students they need to demonstrate. Task writers will often ask a question in such a way that it triggers a single example or piece of supporting information, but then they turn around and characterize as excellent only those responses that have more substance than cued for. This problem can be alleviated in two ways. First, if task writers determine that an excellent response should contain a certain number of examples or supporting ideas, they should make that clear. For example, students may be asked to "Identify three natural resources that are becoming more scarce and tell of two uses that people will have for those resources in the 21st century." Task writers should be cautioned, however, that there is no magic to a certain number of response elements—they simply reflect what a given community of test developers have ascertained to demonstrate accurate and thorough knowledge. More often, the better route may be to cue students to develop ideas rather than to provide a specific number of response elements ("Identify natural resources that are becoming scarce and, using specific examples, explain how people will use those resources in the 21st century"). Responses can then be scored for accuracy and thoroughness, with partial responses being those which are either partially complete or only partially correct.

There is more leeway, in this alternative, because students can provide multiple examples to support an idea or sufficiently elaborate on a single example to earn full credit for that response.

8. **Have you been sensitive to alternative meanings that can trigger legitimate, though unintended, responses?**

One rule of thumb that task writers should remember is that if there is more than one possible meaning for a word, some students will respond based on the alternative meaning. The unintended effect of multiple meanings is evident, for example, in one reading activity that was part of a sample task dealing with two texts on the rain forest. After reading *The Great Kapok Tree*, students were asked "What is one common problem which the animals who lived in the Kapok tree had?" Task writers intended students to read "common" as "shared," but instead, saw responses that dealt with "typical" or "everyday" problems.

The need to consider possible alternative meanings is especially important when developing activities for young students, many of whom interpret questions very literally. For example, when recently asked "If you heated all the corn kernels in a box of popcorn, would they all pop? Explain why or why not," almost all 3rd graders answered, "No, because you need to take the packets out of the box first," or "No, because you're supposed to microwave the paper packet." Even those among us who pride ourselves on having a particular gift for "thinking like a kid" never anticipated such responses.

9. **Have you cued clearly when both similarities and differences are essential to a response or when students legitimately may focus on *either* that which is unique *or* that which is shared?**

Ambiguity of test developers' intentions also comes into play sometimes when activities involve making comparisons. According to most schema of thinking skills, comparison subsumes contrast, and when asking students to make a comparison teachers anticipate that both similarities and differences will be identified. When developing performance tasks, however, the *opportunity* to address both must be clearly distinguished from the *requirement* to do so, and the suitability of these alternatives must be carefully considered. It is fundamentally unfair to ask how a student is both like and unlike a character in a story, for example, if dissimilarities might be all a student could legitimately identify. Particularly when using graphic organizers such as a Venn diagram or comparison matrix in assessments, some educators tend to equate success with completion of the graphic display, although the topic being addressed may not lend itself to such completion. To develop valid, scorable performance assessment activities eliciting comparison, or any other complex thinking process, we must be sure that options are not confused with expectations.

10. **If the *explicit* identification of relationships among ideas, positions, or texts is desired, have you cued for it?**

Unless explicit connections are cued for, students' ideas may only be implied from their responses. Particularly in contexts in which they have all read the same text, students may not recognize that the test-taking situation demands greater explicitness of referents. Thus, we may anticipate that when asked "How are you like the protagonist," some students may only write, "I am a middle school student but I'm tall for my age," instead of providing explicit comparison (e.g., "Like John, I'm in middle school but while he is a runt, I am tall for my age"). This is one situation in which asking for information to be presented in a graphic organizer is quite appropriate; the activity format itself will encourage students to specify the dimensions for comparison and to be more explicit and thorough in their responses.

11. **Have you used the words "an" or "one" instead of "most" or "best" when you anticipate that there is indeed more than one correct response possible?**

Although we may indeed believe that a particular answer is better than some other, when questions are open-ended we must be careful not to send a different signal. Some students, particularly bright ones, sometimes get hung up on the decision-making process involved in selecting the "best" or "right" answer. The frustration triggered by this process may even cause them to leave the activity blank. Simply changing the language to signal choice or multiple possibilities alleviates this problem. Thus, for example, asking "What is one significant contribution your senator has made in recent years?" is better than "What is the most significant contribution your senator has made?"

12. **Consider carefully your use of metacognitive cues like "why" and "how" and be as explicit as possible when cuing for what you want.**

As noted earlier, students, particularly young ones, tend to be literalists. We also know that many students, if not all, do indeed want to please us. (Is there a teacher out there who has not heard a student ask, "Is this what you want?" or "Is this right?") Their earnest desire to answer our questions sometimes leads to quite unexpected results. For example, many students responded to the question "How do you think these groups meet their needs and wants" with "by working together." Not a bad response, except that the task writers were seeking an articulation of process. In that case, what they should have asked was "What steps or what actions should the groups have taken to meet their needs and wants?" I sometimes suggest to task writers that they be as graphic as possible, asking young children to "walk back through the steps they took in their mind to solve the problem" or "explain so somebody could understand the steps you went though in your mind."

The cue "why," can be as troublesome as the cue "how"—particularly for young children. When, after designing several symmetrical forms, students were asked to write a note to a teacher explaining why the forms were symmetrical, they answered "because you told me to make them that way." We do recognize that the inappropriate response to this cue sometimes reflects students' lack of familiarity with the discourse conventions of disciplines like mathematics and science. In those fields, explaining how or why requires support based on data; personal opinions or impressions may be better left to creative or nonacademic discourse. Until students have greater familiarity with those conventions, however, we would not be remiss in providing more clarity to the students who participate in the performance assessments we develop.

## Final Considerations

Among those involved in performance assessment task development, there is some difference of opinion about the degree of scaffolding, or guidance provided through format, examples, and cuing, that is needed or appropriate. Typically, a formative assessment—one intended to provide information in a more informal context on "where are we now"—will be more heavily scaffolded. A summative assessment—one intended to evaluate progress at key points in time and to serve accountability purposes—might arguably include less cuing, since students may be expected to develop their ideas or to be explicit, for example, without specific instructions to do so. Any effective performance assessment task, regardless of its purpose or use, must be presented in such a way that what is valued—and will be evaluated—in a student response is as clear as possible.

Regardless of the nature and purpose of a particular performance assessment, those people involved in task development and use need to scrutinize tasks more carefully in light of the possible, not just intended, response behaviors they may trigger. Applying the lessons learned from scoring to the development of performance assessment tasks will ensure more valid tasks and yield more meaningful results. Furthermore, from educators' increased sensitivity to the impact of the language and format of stimuli on students' performances can come not only better assessment, but also better day-to-day instruction and improved student learning—those goals which have been the impetus, after all, for exploring new ways of eliciting and describing what students understand and can do.

# V-6

# Creating Tests Worth Taking

GRANT WIGGINS
CENTER ON LEARNING, ASSESSMENT, AND SCHOOL STRUCTURE

Copyright © 1992 Center on Learning, Assessment, and School Strucure. First published in *Educational Leadership* 49, 8 (May 1992): 26–33. Reprinted with permission.

# Creating Tests Worth Taking

Grant Wiggins

*The Director of Research at CLASS provides questions, criteria, and suggestions for test designers who want to engage students as well as evaluate their performance.*

GRANT WIGGINS

Should a test be enticing? I think so. And should tests more often be authentic simulations of how knowledge is tested in adult work and civic settings? Many of us believe so. "Performance assessment" calls upon test makers to be creative designers then, not just technicians.

In performance assessment the design issues resemble those facing the architect. There is ground to be covered (the syllabus), there are the logistics of making the design fit the site (making large-scale assessment work in the school), and there are building codes (psychometric norms) and town elders (school board members and district testing directors) to worry about. But designers have typically avoided another, more basic obligation: the need to serve the users — in this case, students and teachers. The clients must "own" the design; form must follow function. The more the tasks (like the house) fit seamlessly with both the environment and the client's aspirations, the better the design and the result.

In this article I offer some proven design tips, tools, and criteria for fashioning assessment tasks that are more enticing, feasible, and defensible — tests worth taking.

## Questions and Criteria

Designers of performance assessments should use the following key questions as a tool to guide the design process:

- What kinds of essential tasks, achievements, habits of mind, or other valued "masteries" are falling through the cracks of conventional tests?
- What are the core performances, roles, or situations that all students should encounter and be expected to master?
- What are the most salient and insightful discriminators in judging actual performances?
- What does genuine mastery of each proposed assessment task look like? Do we have credible and appropriate exemplars to anchor our scoring system? Have we justified standards so they are more than local norms?
- Are the test's necessary constraints — imposed on help available from others, access to resources, time to revise, test secrecy, prior knowledge of standards — authentic?
- Do our assessment tasks have sufficient depth and breadth to allow valid generalizations about overall student competence?
- Have we ensured that the test will not be corrupted by well-intentioned judges of student work?

- Who are the audiences for assessment information, and how should assessment be designed, conducted, and reported to accommodate the needs of each audience? When are audit-tests appropriate and inappropriate?

These questions can be summarized and reframed to produce eight basic design criteria:

1. Assessment tasks should be, whenever possible, authentic and meaningful — worth mastering.
2. The set of tasks should be a valid sample from which apt generalizations about overall performance of complex capacities can be made.
3. The scoring criteria should be authentic, with points awarded or taken off for essential successes and errors, not for what is easy to count or observe.
4. The performance standards that anchor the scoring should be genuine benchmarks, not arbitrary cut scores or provincial school norms.
5. The context of the problems should be rich, realistic, and enticing — with the inevitable constraints on access to time, resources, and advance knowledge of the tasks and standards appropriately minimized.
6. The tasks should be validated.
7. The scoring should be feasible and reliable.
8. Assessment results should be reported and used so that *all* customers for the data are satisfied.

The suggestions and observations that follow offer further assistance to would-be designers.

### Choosing What to Test

*Choose exit outcomes or areas of the curriculum that now tend to fall through the cracks of conventional testing.* Typical tests, even demanding ones, tend to overassess student "knowledge" and underassess student "know-how with knowledge" — that is, intellectual performance. Auditing local tests with Bloom's taxonomy as criteria, for example, shows that synthesis is infrequently assessed at present, and is *inherently resistant* to assessment by multiple-choice tests because it requires "production of a unique communication" that bears the stamp of the student.[1]

**Typical tests, even demanding ones, tend to overassess student "knowledge" and underassess student "know-how with knowledge."**

Faculties should also consider their institutional "customers." What kinds of tasks must our *former* students master? Here, for example, is a question from a freshman final exam in European history at a prestigious college; it suggests how even our better students are often ill-prepared for real intellectual tasks:

> Imagine yourself Karl Marx, living half a century later. Write a brief evaluation of the programs of the Fabian socialists and the American reformers such as T. Roosevelt to present to the Socialist International.

*Think of the knowledge to be tested as a tool for fashioning a performance or product.* Successful task design requires making the essential material of a course a *necessary means* to a successful performance *end*. Example: a 5th grade teacher assesses geography knowledge by having his students devise a complete itinerary, map, and travel packet for their favorite rock group's world tour, within certain budget, logistical, cultural, and demographic restrictions.

Another example: students are asked to design a museum exhibit around a theme studied in a history course, selecting from many real or facsimile artifacts; required to justify what is both included and excluded in the exhibit; and must seek funding from a "foundation" of teachers and peers for the exhibit.

We want to know: Can the student use knowledge and resources effectively, *to achieve a desired effect*? This is the question Bloom and his colleagues argued was at the heart of synthesis. These tasks should only be judged well done to the extent that the content is well used.

### Designing the Tasks

*Contextualize the task.* The aim is to invent an authentic simulation, and like all simulations, case studies, or experiential exercises, the task must be rich in contextual detail. A context is rich if it supports multiple approaches, styles, and solutions and requires good judgments in achieving an effective result. One must please a real audience, make a design actually work, or achieve an aesthetic effect that causes pride or dismay in the result.

The test may be a contrivance, but it needn't *feel* like one.[2] Consider professional training and testing. Doctors and pilots in training confront situations that replicate the challenges to be later faced. Business and law students learn by the case method, fully immersed in the facts of real past

**All real-world performers know the target and the standards, not just their task in advance.**

cases. A context is realistic to the extent that we so accept the premises, constraints, and "feel" of the challenge that our desire to master it makes us lose sight of the extrinsic factors and motives at stake — namely that someone is evaluating us. In just this way, for example, putting out a school newspaper for a journalism course doesn't feel contrived.

Here's an example of how a teacher's attempt to design a performance task evolved as a concern for context was introduced. The original task, in a global studies course, required students to design a trip to China or Japan. But what kind of trip? For what customers? With what constraints of budget or time? The teacher then refined the task so that each student had a $10,000 budget for designing a month-long, cultural-exchange trip for students their age. Still, the purpose is too abstract. What must the tour designers accomplish? Are they trying to design a tour in the abstract or really attract tour-takers? The students were finally charged to be travel agents who develop an extensive brochure and research the cost and logistical information using a computer reservations system.

There is no such thing as performance-in-general. To understand what *kind* and *precision* of answer fits the problem at hand, the student needs contextual detail: it clarifies the desired result, hence the criteria and standards. Too many measurement tasks have an acceptable margin of error that is arbitrary. Are we measuring body temperature or roasts in the oven? It matters. The task's standard of performance (desired precision or quality of product) should be apparent. In fact, an important oversight by the global studies teacher was her failure to give the students model tour brochures.[3]

*Aim to design "meaningful" tasks — not the same as "immediately relevant or practical" tasks.* An assessment task will be meaningful to the extent that it provokes thought and thus engages the student's interest. But a task can be engaging without being of apparent, immediate usefulness. Whether it be mysteries, debates, mock trials, putting on plays — or, for that matter, Nintendo — students clearly respond to "irrelevant" but real challenges. What do such tasks have in common? Designers need to conduct better empirical studies to discover the tasks that tap those twin intellectual needs: *our urge for efficacy and our need for meaningful connections.*

This caution about meaning vs. relevance is particularly warranted to avoid turning important theoretical problems into crude utilitarian ones. Many genuine problems do not have obvious practical value, but they nonetheless evoke interest and provide insight into student abilities. Consider two such problems, one in geometry and one in history/English:

> We all know the Pythagorean theorem: $A^2 + B^2 = C^2$; but does it have to be a square that we draw on each leg? Suppose we drew the same shape on each leg; would the areas on A and B add up to the area on C? Find other shapes that make the equation work, too, and try to derive a more general formula of the theorem.[4]
>
> You and your colleagues (groups of 3 or 4) have been asked to submit a proposal to write a U.S. history textbook for middle school students. The publishers demand two things: that the book hit the most important things, and that it be interesting to students. Because of your expertise in 18th-century American history, you will provide them a draft chapter on the 18th century, up to but not including the Revolution, and "field tested" on some middle school students. They also ask that you fill in an "importance" chart with your response to these questions: (1) Which event, person, or idea is most important in this time period, and why? (2) Which of three sources of history — ideas, people, events — is most important? You will be expected to justify your choices of "most important" and to demonstrate that the target population will likely be interested in your book.

*Design performances, not drills. A test of many items (a drill) is not a test of knowledge in use.* "Performance" is not just doing simplistic tasks that cue us for the desired bit of knowledge. It entails "putting it all together" with good judgment; good judgment cannot be tested through isolated, pat drills. As one teacher put it to me a few years ago: "The trouble with kids today is that they don't know what to do when they don't know what to do." She is right — and a prime reason is that tests rarely put students in an authentic performance situation, where *thinking*, not just an obvious bit of knowledge, is required.

The designer's aim, then, is to avoid inventing a new round of (this time, hands-on) isolated items. Rather, we should consider the difference between drilled ability vs. performance ability and ask: *What is the equivalent of the game or recital in each subject matter?* What does the "doing" of mathematics, history, science, art, language use, and so forth, look and feel like in context? What are the projects and other kinds of synthesizing tasks performed all the time by professionals, consumers, or citizens that can be adapted to school use?

Such tasks are always "higher-order," and we would do well to use Lauren Resnick's criteria in our search for better-designed assessments. Higher-order thinking

- is *nonalgorithmic* — that is, the path of action is not fully specified in advance;
- is *complex*, with the total path not visible from any single vantage point;
- *often yields multiple solutions*, each with costs and benefits;
- involves *nuanced judgment* and interpretation;
- involves the *application of multiple criteria*, which sometimes conflict with one another;
- often involves *uncertainty*, because not everything that bears on the task is known;
- involves *self-regulation* of the thinking process, rather than coaching at each step;
- involves *imposing meaning*, finding structure in apparent disorder;
- is *effortful*, with considerable mental work involved.[illegible]

It may help to think of this problem as the search for larger, more interrelated but complex chunks of content to build tasks around. What, for example, might be 8 to 10 important performance tasks in a subject that effectively and efficiently "map" the essential content? Vocational programs usually grapple well with this problem by casting the course objectives as a set of increasingly complex tasks to be mastered, in which the student in the last task(s) must literally put it all together, for example, build a house in carpentry.

Grant Wiggins

*Refine the tasks you design by building them backwards from the models and scoring criteria.* A complex task is not a vague task, with the objective or specifications unknown. All real-world performers know the target and the standards, not just their task in advance; such knowledge guides their training and rehearsals. Students should never have to wonder "Is this right?" "Am I finished?" "How am I doing?" "Is this what you want?" In a "real" problem the task is ill-structured but well-defined: the goal, specifications, or desired effect is known, but it is not obvious how to meet it. Knowing the *requirements of task mastery* — the "specs" — means the student must be habituated by testing to think of mastery as control over the *knowable* essentials, not as calculated cramming and good guesses. This requires providing the student with scoring criteria and models of excellent performance or production as part of instruction. (Think of diving and debate.) Such practice is the norm throughout Carleton, Ontario, where students work from "exemplar booklets" to practice grading student work — in the same way now reserved for judges in our assessments.

"What does mastery at the task look like? What will we be able to properly infer from the collected student work?" These become the *key* questions to ask in the challenge of taking a basic idea and making a valid performance-assessment task out of it (as opposed to an instructional task). The questions properly focus on judging anticipated results and move away from design that produces merely pleasant or interesting work.

## Scoring Considerations

*Score what is most important for doing an effective job, not what is*

Grant Wiggins

*easiest to score.* The scoring rubrics should represent generalizations about the traits found in an array of actual performances. But too often we resort to scoring what is easiest — or least controversial — to observe. A fine task can be rendered inauthentic by such bogus criteria.

Two key questions for setting up a scoring system therefore are: "What are the most salient characteristics of each level or quality of response?" and "What are the errors that are most *justifiable* for use in lowering a score?" Obvious successes and errors (such as those that relate to spelling or computation) are not necessarily the most accurate indicators of mastery or its absence.[6] Too many essay scoring systems reward students for including merely *more* arguments or examples; quantity is not quality, and we teach a bad lesson by such scoring practices.

When possible, scoring criteria should rely on descriptive language, not evaluative and/or comparative language such as "excellent" or "fair." Judges should know specifically where in performance to look and what to look for. The ACTFL foreign language proficiency guidelines and the Victoria, Australia, "Literacy Profiles" are perhaps the best examples available of such empirically grounded criteria.[7] Teachers may also want to have students analyze a task and help devise the scoring system. This builds ownership of the evaluation, makes it clear that judgments need not be arbitrary, and makes it possible to hold students to higher standards because criteria are clear and reasonable.

*"Benchmark" the standards for performance to ensure that your scoring standards are wisely chosen and suited to wider-world or next-level demands.* Standard-setting for performance involves selecting exemplary samples of performance or production. The challenge is to avoid using local age-grade norms; the solution is to equate our exit-level standards to wider-world entry-level standards at desirable colleges or professions. That advice, of course, begs a more fundamental question: Whose view of excellence should count? It is at least prudent to equate local standards of scoring to some credible wider-world or next-level standard — something routinely done in the performing arts, athletics, and vocational education.[8] *And, every so often, refer to next-level standards when scoring the work of younger students.* (I believe Illinois was the first state to assess both 6th and 8th grade writing samples against 8th grade exemplars, for instance.)

### Administering the Assessments

*Since constraints always exist in testing, make them as authentic as possible.* The question is not "Should there be constraints in testing?" but rather "When are constraints authentic, and when are they inauthentic?" It is often a matter of degree, but the principle needs to be maintained and defended.

Constraints facing the designer of authentic assessment tasks typically involve access or restrictions to the following resources: (1) time (including time to prepare, rethink, and revise), (2) reference material, (3) other people (including access to peers, experts, the test designer, and/or the judge), and (4) prior knowledge of the tasks and how they will be assessed (the issue of test security). The question then becomes: What are *appropriate* limits on the availability of these resources?

Traditional testing, because it involves indirect proxies for performance, requires numerous inauthentic constraints to preserve validity. The validity of most multiple-choice tests, for example, is compromised if questions are known in advance or if reference material can be consulted during the test. These habits of administration run deep; they seem obviously required. But what of the validity issues raised by denying students access to basic resources? Just what is being tested when the student cannot predict the historical periods or books that will be assessed, or cannot consult resources while writing?

We need not keep textbooks and other materials from students if the task is genuinely authentic. For example, in many of Connecticut's performance tasks in mathematics, the key formulas are given to the student as background to the problem. And why not allow the student to bring notes to the exam? A physics teacher I know allows students to bring an index card to the exam with anything on it; the card often reveals more about the student's knowledge than the exam answers!

Too little time for performing is not always the key issue either. Is the limiting of the test to *one sitting* authentic? If writing is indeed revision, for example, why not allow writing assessment to occur over three days, with each draft graded? Many districts now do so, including Jefferson County, Kentucky, and Cherry Creek, Colorado.[9]

I am not arguing that the student should have unlimited time and access in testing.[10] Let us ask: What kinds of constraints authentically simulate or replicate the constraints and opportunities facing the performer in context? What kinds of constraints tend to bring out the best in apprentice performers and producers?

*Develop a written, thorough protocol that details how the task should be administered — especially so judges will know the proper limits of their interventions to student acts, comments, or questions.* It is incredibly easy to invalidate performance assessment by varying the instructions, the amount of assistance provided, and the depth of responses given to inevitable student questions. Determining beforehand what is acceptable response and intervention by adults is essential; test administrators must receive standard verbal responses for delicate situations, confusions, or problems that arise.

And don't forget that kids can do the darndest things with directions that aren't thought through. In a hands-on science experiment that asked whether "the sun" heated up different colored liquids at different rates, a student did not use the heat lamp provided, moved all his equipment to the window, saw it was a cloudy day, and wrote "no."

*Make the tasks maximally self-sustaining and the record-keeping obligation mostly the student's.* Many educators who have never seen large-scale performance assessment cannot fathom how all students can be efficiently and effectively assessed. But they assume that the teacher will have to guide activity every step of the way and record massive amounts of information simultaneously. Thoughtful preparation, designed to make the assessment self-running, frees the teacher to be a perceptive judge.

### Creating a Tool Kit

*Develop a districtwide "tool kit" of exemplary tasks, task templates, and design criteria for assessment tasks.* Not all of us are good designers, but why should we have to be? Teachers can help their colleagues by providing a sampler of tasks and task templates. Kentucky has done this at the statewide level, providing dozens of tasks and task ideas to teachers as part of the new state performance-based assessment system. We should consider including not only current examples of model assessment tasks, but traditional performance-based challenges such as debates, treasure hunts, mysteries, design competitions, historical reenactments, science fairs, Odyssey of the Mind tasks, Scout Merit Badges, student-run banks and stores, and so forth.

The mathematics performance assessment team of the Connecticut Department of Education has identified the following types of problems as central to its work:

- Given data on graphs, write a story that represents the data or graph.
- Given headlines or claims with background data, explain whether or not the claims are reasonable.
- Given student work containing common errors, write a response to the student.
- Given equations or number facts, write a problem that the equations or facts could solve.
- Given trends or sample data, make and justify predictions.
- Given consumer- or job-related buying, selling, or measuring situations, solve a problem.
- Given multiple or competing interpretations of given data, justify each interpretation.

Job roles provide ample opportunities for task designers to create simulations. Here are some suggestions.

- Museum curator: design museum exhibits; compete for "grant" money.
- Engineer or surveyor: bid and meet specs for largest-volume oil container; build a working roller

**Assessment design is like software design: you can *never* accurately and fully anticipate the naive user's response.**

**Good teaching is inseparable from good assessing.**

coaster; map or survey a region around school or in the building.

- Ad agency director: design advertising campaigns, book jackets, or blurbs for books read in class.
- Psychologist/sociologist: conduct surveys, perform statistical analyses, graph results, write newspaper articles on the meaning of results.
- Archaeologist: determine the culture or time frame of a mystery artifact or person.
- Newspaper editor and writer: write articles and editorials set in the studied historical time.
- Policy analyst: predict the future in a country being studied.
- Product designer: conduct research, design ad campaign, present proposal to panel.
- Job interviewee: present portfolio and try to get "hired" for a specific job related to skills of current course (interview conducted by other students or teacher).
- Expert witness to Congress: testify on behalf of or against advertising claims, regulation of children's TV, or current policy issue.
- Commercial designer: Propose artwork for public buildings.

### Piloting and Reporting

*Always pilot some or all of the test.* Assessment design is like software design: one can *never* accurately and fully anticipate the naive user's response. A first design may not fit the purpose or maximally evoke the desired knowledge; a prompt might result in irrelevant responses that are nonetheless appropriate or reasonable to the student; the logistical constraints of a context can turn out to be more daunting than anticipated; the judges may be too self-interested in the results or insufficiently trained. A pilot is the only way to find out, even if it involves only a tiny sample of performers. And the de-bugging requires a *naive* "guinea pig" — a teacher from a different subject or a few students — if the hidden problems in the goal, directions, or procedures are to be found.

*You are what you report: Make sure that your report cards, transcripts, and district accountability reports relate achievement and progress to essential performance tasks and exit-level standards.* Few transcripts reflect achievement in reference to outcomes. They tend to certify that tests were passed on each isolated packet of content instead of documenting what the student can do and to what level of performance. Further, a one-shot test cannot validly assess many important capacities, as the phrases "habits of mind" or "consistency of performance" suggest. Grading and reporting thus need to move toward scoring that provides a "progress" measure — that is, work scored against exit-level performance standards. And no worthy performance is reducible to one aggregate score. Every student ought to have the equivalent of a baseball card — many different kinds of abilities measured and a brief narrative report — if we are seriously interested in accurately documenting and improving complex performance.

### Assessment's Role in School Reform

An underlying premise of this kind of assessment reveals why I believe that assessment reform is the Trojan horse of real school reform. We badly need better definitions of mastery or *understanding* to guide assessment design, curriculum design, and teacher job descriptions and performance appraisal. Circling "correct" answers to problems only test makers care about is not "knowing," nor is it the aim of teaching. Authentic tests provide a stimulating challenge instead of an onerous obligation.

Perhaps more important for school restructuring is the need to build local educator capacity and interest in quality assessment.[11] Genuine faculty empowerment is impossible without deep ownership of local standards and measures. Farming all these problems out to distant "experts" is a grave mistake — one rarely made in any other country. Good teaching is inseparable from good assessing. It may well be that experts can design more rigorous tests, and that correlational/predictive validities exist in standardized tests. But schooling we can be proud of and held genuinely accountable for demands more locally useful, authentic, and enticing assessments.

---

'Bloom, (1954), pp. 163, 175. Serious would-be test designers would do well to reread the *text* of the taxonomy, not just the

Appendix/list, as well as the follow-up handbook developed by Bloom, Madaus, and Hastings, (1981).

[2] "The student should [have] freedom from excessive tension . . . be made to feel that the product of his efforts need not conform to the views of the instructor . . . [and] have considerable freedom of activity . . . [including] freedom to determine the materials or other elements that go into the final product." In Bloom, (1954), p. 173.

[3] See Linn, Baker, and Dunbar, (1991), for further discussion of validity design issues.

[4] I have watched half a dozen classes immerse themselves in this problem and beg to continue when time ran out.

[5] From Resnick (1987).

[6] Describing key errors and using them in the rubric is a *very different* matter than building them into test answers as "distractors."

[7] A related issue that emerges in designing rubrics (and thus far unaddressed by measurement experts) is the difference between the degree of difficulty of the task and the desired quality of the performance — a distinction made in New York's music performance assessments.

[8] See Wiggins (1991).

[9] Yes, yes, I know the issue is *really* one of cheating. Let the teacher "sign off" on the papers, then, certify authorship, as they do in Australia and now in Vermont.

[10] Though many New York State tests do allow the student what amounts to unlimited time — all day — given the shortness of the test. And certifiably learning disabled students are allowed unlimited time on the SATs as well as many state achievement tests.

[11] See Stiggins (1991).

## References and Readings

Bloom, B. S., ed. (1954). *Taxonomy of Educational Objectives: The Classification of Educational Goals; Handbook I: The Cognitive Domain.* New York: Longman Publishers.

Bloom, B. S., G. F. Madaus, and J. T. Hastings. (1981). *Evaluation to Improve Learning.* New York: McGraw-Hill. [A major revision of *Handbook on Formative and Summative Evaluation of Student Learning.* (1971). McGraw-Hill].

Linn, R. L., E. L. Baker, and S. B. Dunbar. (November 1991). "Complex, Performance-Based Assessment: Expectations and Validation Criteria." *Educational Researcher* 20, 8: 15-21.

Mitchell, R. (1992). *Testing for Learning: How New Approaches to Evaluation Can Improve American Schools.* New York: The Free Press.

National Council of Teachers of Mathematics. (Forthcoming). *Mathematics Assessment: Myths, Models, Good Questions, and Practical Suggestions.* Reston, Va.: NCTM.

Resnick, L. (1987). *Educ* [illegible] *d Learning to Think.* Was [illegible] D.C.: National Academy Press.

Schwartz, J. L., and K. A. Viator, eds. (1990). "The Prices of Secrecy: The Social, Intellectual, and Psychological Costs of Testing in America." A Report to the Ford Foundation. Education Technology Center, Cambridge, Mass.: Harvard Graduate School of Education.

Stiggins, R. (March 1991). "Assessment Literacy." *Phi Delta Kappan* 72, 7: 534-539.

Victoria Ministry of Education. (1990). *Literacy Profiles: Handbook.* Victoria, Australia. [Distributed in the U. S. by TASA, Brewster, N.Y.]

Wiggins, G. (February 1991). "Standards, Not Standardization: Evoking Quality Student Work." *Educational Leadership* 48, 5:18-25.

Wiggins, G. (June 1990). "Finding Time." *Basic Education* 34, 10.

Wiggins, G. (May 1989). "A True Test: Toward More Authentic and Equitable Assessment." *Phi Delta Kappan* 70, 9: 703-713.

Wiggins, G. (April 1989). "Teaching to the (Authentic) Test." *Educational Leadership* 46, 7: 41-47.

Copyright ©1992, Center on Learning, Assessment, and School Structure.

# V-7

# What We've Learned About Assessing Hands-On Science

RICHARD J. SHAVELSON AND GAIL P. BAXTER
UNIVERSITY OF CALIFORNIA, SANTA BARBARA

First published in *Educational Leadership* 49, 8 (May 1992): 20–25. Reprinted with permission from the authors.

# What We've Learned About Assessing Hands-On Science

*Assessing scientific inquiry is more complex than political rhetoric pushing performance tasks indicates, this team of scientists found. And, unless carefully crafted and blended into science instruction, assessments alone are unlikely to boost achievement.*

RICHARD J. SHAVELSON AND GAIL P. BAXTER

Over the past three years, our team of researchers, scientists, and science teachers at the University of California, Santa Barbara, the California Institute of Technology, and the Pasadena Unified School District has sought to create assessments that support good science teaching at the elementary level (Baxter et al. in press; Shavelson et al. 1991). In particular, our goal has been to develop activities that permit students to pursue an experimental inquiry focusing on process skills (such as observing and inferring) and construction of new knowledge (such as understanding the effects of insulators on electric current). In our definition, assessments consistent with good teaching invite students to perform concrete, meaningful tasks such as a laboratory experiment to determine, for example, how certain kinds of insects respond to changes in environment. Scoring of performance focuses on the reasonableness of the procedure used to carry out the investigation, not just on the "right answer."

In our work, we assumed that the ideal assessment would be direct observation of a student pursuing a scientific inquiry with laboratory equipment and materials. This observation would be made by scientists and science teachers trained to score performance in real time. That is, the ideal was predicated on the assumption of the symmetry of teaching and testing: an ideal assessment would be a good teaching activity and, indeed, might even serve as a teaching activity when not used for assessment.

However, observations of individual student performance are costly, time consuming, and difficult to obtain. With the ideal performance assessment as a benchmark, we developed and evaluated alternatives (or surrogates) to the benchmark. They were, in order of decreasing verisimilitude, (1) lab notebooks in which students recorded their procedures and conclusions; (2) computer simulations of the hands-on investigations; (3) short-answer paper-and-pencil problems in planning, analyzing, and/or interpreting experiments; and (4) multiple-choice items developed from observations of students conducting hands-on investigations. Finally, we compared the benchmark and surrogate assessments to a traditional multiple-choice science achievement test, the Comprehensive Test of Basic Skills (CTBS).

We developed and collected data using three hands-on investigations:

- Paper Towels — Given laboratory equipment, conduct an experiment to determine which of three different paper towels soaks up the most water.
- Electric Mysteries — Given six "mystery boxes," determine their contents by connecting electric circuits to them.
- Bugs — Determine sow bugs' preferences for various environments (for example, dark or light, dry or wet).

The performance of more than 300 students, some experienced in hands-on science and some who had received minimal science instruction from a textbook on health, was observed and scored in real time by science educators. In addition, all students completed corresponding notebooks, computer simulations, paper-and-pencil measures, and the CTBS.

Four questions guided our research: (1) Could reliable measures of hands-on performance and of surrogate assessments be developed? We wanted these measures to permit a wide variety of student responses found when doing science. We also wanted to develop a method to score performance that captured the diversity of procedures and put them on a common scale. (2) Could the performance of students with different instructional experiences (hands-on vs. textbook) be distinguished? We expected students experienced with hands-on science to perform better on the benchmark and closest alternatives than students in a textbook curriculum program. (3) Do the performance assessments provide information about student achievement not available from traditional multiple-choice science tests? If not, perhaps nothing new had been developed. (4) Do the surrogate assessments capture the information gained from the benchmark? If so, then dollars and time can be saved in administration and scoring.

### Hands-On Investigations

Students conducted the three investigations in approximately 1 1/2 hours while being observed by a scientist or science educator trained to score student performance.

*Paper Towels.* Students used a laboratory setup to determine which of three paper towels held the most and least water. Students were told that they could use all or some of the equipment, whatever they needed. A scheme was developed to score the diversity of procedures used to carry out the investigation on a common scale (see fig. 1). An outstanding investigation completely saturated each towel, determined the amount of water each held by a method that was consistent with the way the towel was wetted, and the entire procedure was done carefully. For example, a student might saturate the towel in the pitcher of water and weigh it in the scale, carefully removing the excess water in the scale after weighing each towel. Carelessness, inconsistencies in the method of wetting the towel and measuring the results, incomplete saturation, and irrelevant methods resulted in less than outstanding scores. The scoring scheme identified the procedure used and could thereby characterize performance in terms of both processes and outcomes. Moreover, several different performances

Jerry Pine

*High-quality assessments throughout a course are vital if teachers are to accurately appraise performance.*

FIGURE 1

**PAPER TOWELS INVESTIGATION—HANDS-ON SCORE FORM**

Student____________ Observer____________ Score ________ Script________

1. **Method**
   A. Container
   Pour water in/put towel in
   Put towel in/pour water in
   1 pitcher or 3 beakers/glasses
   B. Drops
   C. Tray (surface)
   towel on tray/pour water on
   pour water on tray/wipe up

2. **Saturation** A. Yes B. No C. Controlled

3. **Determine Result**
   A. Weigh towel
   B. Squeeze towel/measure water (weight or volume)
   C. Measure water in/out
   D. Time to soak up water
   E. No measurement
   F. Count # drops until saturated
   G. See how far drops spread out
   H. Other __________

BEST COPY AVAILABLE

4. **Care in Measuring** Yes No

5. **Correct Result** Yes No

| Grade | Method | Saturate | Determine Result | Care in Measuring | Correct Answer |
|---|---|---|---|---|---|
| A | Yes | Yes | Yes | Yes | Yes |
| B | Yes | Yes | Yes | No | Yes/No |
| C | Yes | Yes/Controlled | Error | | Yes/No |
| D | Yes | No | Missing | | Yes/No |
| F | ---------- | ---------- | No Attempt | ---------- | ---------- |

could result in the same letter grade.

*Bugs.* Students used laboratory apparatus to determine the preferences of sow bugs for environments that were light or dark, damp or dry, or some combination of the four. The scoring scheme resembled the one used in the towels investigation.

*Electric Mysteries.* This investigation was a bit different from the others. Students were given batteries, bulbs, and wires and asked to determine the circuit components hidden in a set of mystery boxes (see fig. 2). Performance was scored on the basis of (1) their determination of the contents of each box and (2) the sequence of tests they conducted to determine the contents.

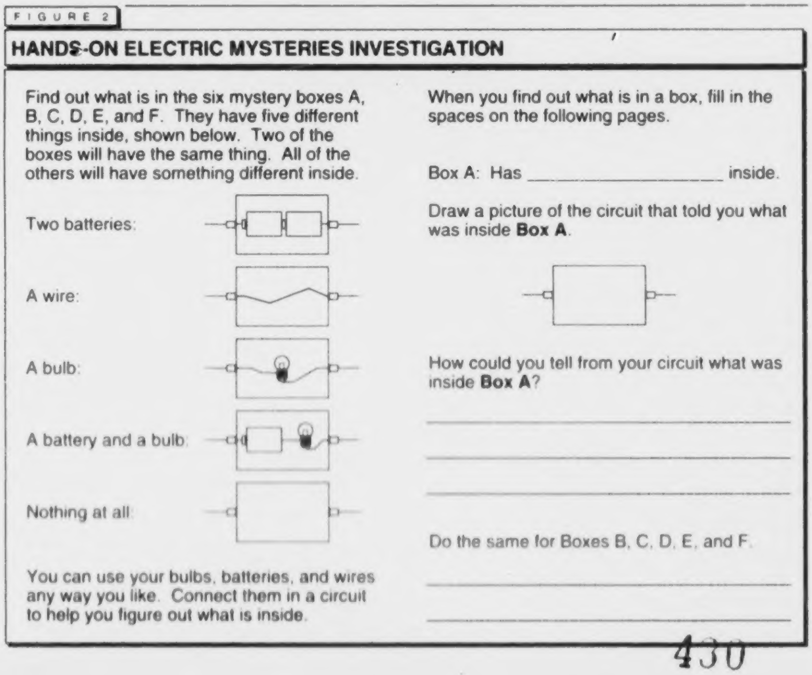
## Notebook Surrogates

For each investigation students were asked to keep notebooks describing their investigation so that a friend could repeat it exactly. Notebooks have several advantages for large-scale assessment. Through the use of notebooks instead of expert observers, many students can be tested with hands-on investigations. Notebooks provide an opportunity for students to express themselves in writing, an important skill in science and a way of integrating curricular areas. And the notebooks can be scored quickly — in about one to two minutes per student. Notebooks, then, preserve much of the hands-on investigation while reducing time and cost of expert observers. They also capture the rather inventive nature of the investigations and ways of reporting on them.

## Computer Simulation Surrogates

We developed our own computer simulation for the Electric Mysteries and Bugs investigations to replicate, as nearly as possible, the hands-on investigations. (The Paper Towels investigation could not be simulated adequately.) For the electric circuits investigation, students used a Macintosh computer with a mouse to connect circuits to the mystery boxes to determine their contents. The intensity of the luminosity of the bulb in a real external circuit was accurately simulated. Students could connect a multitude of circuits if they so desired. Alternatively, they could leave a completed circuit on the screen for comparison. A teacher-directed tutorial prior to the test provided students with instructions on how to record their answers, erase wires, save their work, or look at a previous page of their work on the screen. The computer recorded every move the student made. The bug simulation was constructed similarly, though it was not possible to record every move the student (and the bugs) made.

FIGURE 2



**HANDS-ON ELECTRIC MYSTERIES INVESTIGATION**

Find out what is in the six mystery boxes A, B, C, D, E, and F. They have five different things inside, shown below. Two of the boxes will have the same thing. All of the others will have something different inside.

Two batteries:

A wire:

A bulb:

A battery and a bulb:

Nothing at all:

You can use your bulbs, batteries, and wires any way you like. Connect them in a circuit to help you figure out what is inside.

When you find out what is in a box, fill in the spaces on the following pages.

Box A: Has ________________ inside.

Draw a picture of the circuit that told you what was inside **Box A**.

How could you tell from your circuit what was inside **Box A**?

________________

________________

________________

Do the same for Boxes B, C, D, E, and F.

________________

________________

Computer simulations have a number of desirable properties for assessment. They are less costly and time consuming to administer than hands-on assessments, though development costs are considerable. Students can be tested in groups by a parent or volunteer who has been briefed on how the simulations work. Student performance can be scored quickly and easily using the scoring system developed for the hands-on investigations. In addition, a computer simulation maintains a full record of performance, so that teachers and/or students can review problem-solving processes. Finally, students experiment with the technology, discovering solutions to problems that they might not have found with other types of assessments. In other words, the computer assessment provides a good instructional tool.

## Pencil-and-Paper Surrogates

Short-answer and multiple-choice questions were chosen to parallel, in content, the three hands-on investigations. For the Electric Mysteries short-answer questions, students received a pictorial representation of a problem similar to one encountered during the hands-on investigation. For example, students might be asked how they would determine the contents of a particular mystery box without looking inside. For the Paper Towels and Bugs questions, students were given descriptions of portions of the investigations and questioned about the control of variables, the setup of the experiment, or the best method to use for measuring results.

Multiple-choice questions began much like the short-answer questions. Rather than formulate a response, students chose among four alternatives, all of which were based on observed performance. For example, an Electric Mysteries question presented alternative circuits connected to a box and asked students to indicate which circuit would tell them what was inside the box.

The paper-and-pencil surrogate assessments differ fundamentally from the other surrogates; they do not respond to the actions taken by the students. Even if a paper-and-pencil test provided immediate written feedback on a decision made by a student (Shulman and Elstein 1975), we doubt it would have the same impact as the feedback from the real-life (hands-on) or lifelike (computer) assessments. We may not be able to develop paper-and-pencil surrogates that overcome this limitation.

## Findings

We found the process of creating performance assessments symmetric with good teaching activities to be time consuming, requiring considerable scientific and technological know-how. Development of quality performance assessments requires multiple iterations through a sequence of development, tryouts with students (getting their thoughts and comments), and revision. Short-circuiting this process leads to ill-conceived and poorly constructed assessments. Such assessments are as likely to lead to poor teaching — if teachers teach to the test, and they do (Smith 1991) — as are ill-conceived and poorly executed classroom activities.

**If teachers teach to poorly constructed assessments, these assessments are likely to distort good hands-on science teaching.**

Once the performance assessments were constructed, our research posed four questions about them: (1) Can they provide reliable measurements? (2) Are they sensitive to students' instructional experiences? (3) Do they provide achievement information that differs from traditional measures? (4) Are they interchangeable? Our findings (some good news, some bad news) balance the political rhetoric pushing implementation of performance assessments with a cold reality.

*Reliability.* Raters can reliably evaluate students' hands-on performance on complex tasks in real time. Reliabilities are high enough (above 0.80) that a single rater can provide a reliable score. But task-sampling variability is considerable. Some students perform well on one investigation while others perform well on a different investigation: general "expertise" is more in the mind of the beholder than in performance. To get an accurate picture of individual student science achievement, the student must perform a substantial number of investigations — perhaps between 10 and 20.

*Instructional history.* Performance assessments can distinguish students with different instructional histories. Assessments that are closely linked to a specific domain of knowledge (for example, electric circuits/electricity) are more sensitive to performance differences than more general process assessments (for example, the Paper Towel investigation). But to be sensitive to instructional history, performance assessments must be carefully crafted to measure more than science process. They need to measure the application of both concepts and procedures.

*Relation to multiple-choice tests.* The good news is that performance assessments do *not* duplicate information about student achievement provided by traditional tests (average correlation is about 0.45). They tap somewhat different aspects of achievement. But we are not sure what aspects of achievement multiple-choice tests or performance assessments do and do not tap. Indeed, a combination of indicators (multiple-choice, performance assessments, and others) may be needed to provide a complete picture of achievement.

*Interchangeability of surrogates.* Certain surrogate assessments appear to be interchangeable with their corresponding benchmark. This is especially true of notebooks for student-level assessment. Computer simulations are interchangeable with their corresponding benchmarks if the intent is to estimate *classroom-level* mean performance. But for individual students, measures of science achievement are highly sensitive not only to the investigation used (for example, Bugs vs. Electric Mysteries), but also to the method used to measure performance (for example, observation vs. simulation). Some students' scores depend on the particular investigation

(Electric Mysteries, Paper Towels, Bugs) *and* on the particular method used to assess performance (observation, notebook, computer simulation, paper-and-pencil). Indeed, each combination of investigation and method provides different insight into what students know and can do.

*Conclusions.* A fundamental assumption made by policymakers (for example, Bush 1991) and other education reformers (for example, Wiggins 1989) is that, by changing the nature of the achievement test, teachers who teach to the test will have to change and improve their teaching. Our experience with performance assessments suggests that this assumption is, at best, half true. Teachers will indeed change the way they teach if held accountable by performance assessments. But, without high quality assessments and staff development in quality instruction, they very well may not improve their teaching. Moreover, one-shot, end-of-year tests cannot provide adequate information on individual-level student achievement. Continuous assessment throughout the course of instruction is needed to accurately reflect student science achievement.

### Assessment Development

Performance assessments are very delicate instruments. They need to be carefully crafted, each requiring a specially developed or adapted scoring method. Shortcuts taken in developing these assessments will likely produce poor measuring devices. If these instruments are used to judge the quality of education in classrooms — and they will be used for that purpose — then teachers will teach to the test. If teachers teach to poorly constructed assessments, these assessments are likely to distort good hands-on science teaching.

This conclusion was brought home to us in our observations of someone we consider to be an outstanding hands-on science teacher. In her class, a unit on electricity is taught in a series of lessons. Small groups of students conduct hands-on investigations, not unlike the ones we have developed. Students keep notebooks and draw conclusions based on the outcomes of their experiments and their discussions with other groups of "scientists" in the class.

Excited about hands-on science teaching, our teacher volunteered her class for pilot testing California's new hands-on science assessments. These assessments were constructed under severe time and cost constraints, and consequently involved minimal trials with students. To meet testing time and space constraints, and in recognition of differing curriculums in the state's elementary schools, the assessments were accompanied by detailed directions to students. Students read the instructions, followed explicit procedures, and reported what they found in the spaces provided. The assessments were more like recipes (first do this step, then do this step) than like scientific investigations.

Based on her experience with these assessments, our exemplary teacher began to modify her teaching to correspond to the state's pilot assessments. Rather than open-ended, group problem solving, emphasis was placed on reading instructions and carefully following format, carrying out a set of required procedures, and recording findings in a prespecified format in notebooks. (For example, students were admonished to "be sure to write complete sentences or the state will grade you down.") The essence of *doing science* was becoming one of following procedures. The story ends happily. After we pointed out what was happening, the teacher went back to her "old ways."

### Staff Development

Unless provided the scientific and pedagogical knowledge required for hands-on science teaching, teachers may very well flounder in their attempts to match their teaching to the testing. Once again, an anecdote may bring home this conclusion. One of the teachers whose class was participating in our research knew of the Paper Towels investigation. Teaching to the test, she instructed her students to saturate the towels completely, using the timer to ensure that each towel was saturated for at least 10 minutes — a total of 30 minutes for saturation! In reality, the towels could be saturated in a matter of seconds. This led her students, when tested, to perform in a clearly stylized manner. She had informed the students, perhaps unintentionally, that science is a set of precise steps that must be carried out invariably, regardless of whether they make sense. Her approach was not particularly conducive to scientific exploration. Although the teacher could teach the students how to "do" the experiment, what was missing was an understanding of the essence of doing science.

### Curriculum-Embedded Assessments

To obtain sufficiently large samples of student performance, assessments may need to be taken throughout the academic year. For example, students might receive the Electric Mysteries assessment embedded in the activities composing a unit on electricity. Likewise, assessments would be embedded in three or four other units as well.

The embedded assessments have several desirable characteristics. They provide almost immediate feedback to teachers on their students' performance, and on how this performance compares with that of students in comparison schools. Moreover, the assessments reinforce hands-on teaching and learning. Nowhere is the symmetry between teaching and assessment more apparent than with embedded assessments.

Embedded assessments do not preclude an end-of-year examination. The latter provides both additional information on achievement and an external audit ensuring data credibility to the various audiences interested in educational accountability.

## Will Achievement Improve?

Results of our research suggest that the political rhetoric calling for immediate reform of national, state, and local testing systems far exceeds current technological capability and ignores educational and social consequences. No doubt assessment systems will be changed in the very near future. The politicians will have their day. We suspect that the initial impact will be to change classroom activities and the nature of assessment, possibly embedding assessments in classroom activities. However, without quality instrumentation and extensive staff development, the bottom line — achievement — will probably not change.

The nation may be capable of producing the kind of assessment systems currently envisioned by the rhetoric if the politicians stick to their guns and do not retreat to the usual multiple-choice testing. Politicians need to provide resources for preparing beginning and current teachers for teaching and testing reforms, and for fine-tuning the assessments through research, social debate, and revision. With the symmetry between assessment and teaching firmly established, the bottom line — achievement — may very well improve. □

*Authors' note*: This research was supported by a grant from the National Science Foundation (No. SPA-8751511). The ideas presented reflect those of the authors and not necessarily the NSF.

## References

Baxter, G. P., R. J. Shavelson, S. R. Goldman, and J. Pine. (In press). "Procedure-Based Scoring for Hands-On Science Assessments." *Journal of Educational Measurement.*

Bush, G. W. (1991). *America 2000: An Education Strategy.* Washington, D.C.: U.S. Department of Education.

Shavelson, R. J., G. P. Baxter, J. Pine, and J. Yure. (1991). "Alternative Technologies for Assessing Science Understanding." Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Shulman, L. S., and A. S. Elstein. (1975). "Studies of Problem Solving, Judgment, and Decision Making: Implications for Educational Research." In *Review of Research in Education* 3: 3-42, edited by F. N. Kerlinger.

Smith, M. L. (1991). "Put to the Test: The Effects of External Testing on Teachers." *Educational Researcher* 20, 5: 8-11.

Wiggins, G. (1989). "A True Test: Toward More Authentic and Equitable Assessment." *Phi Delta Kappan* 70, 9: 703-713.

# V-8

# The Development of Performance Assessment Exercises

RICHARD J. STIGGINS
ASSESSMENT TRAINING INSTITUTE

To understand the treatment of the topic of performance assessment exercise design presented in this paper, we must place that treatment in the context of a broader understanding of the overall challenge of devising quality assessments. For example, we must realize that performance assessment represents one of four basic forms of assessment available to educators. The others are selected response assessments, essay assessments, and direct personal communication with the student. All represent viable options. In addition, we must remain aware of the fact that the method of choice in any particular context is a function of the achievement target to be assessed. Targets vary. Sometimes we wish to assess mastery of subject matter knowledge. Other times we attend to the ability to use that knowledge to reason and solve problems. Still other times we seek to assess mastery of skills or the ability to use those skills to create quality products. All such targets are important and may represent the focus of classroom or large-scale assessments. Performance assessment represents a powerful method for assessing reasoning proficiency, the skillfulness of student performance, or the ability to create quality products. These, then, are the targets performance assessment exercises must sample. (For a more complete discussion of the match of methods to targets, see my paper in Section I of this guide.)

We must also remain cognizant of the fact that performance exercises represent only one part of the performance assessment tool. Other parts include the performance criteria and the scoring and recording procedures. To understand exercise development as described herein, we must understand the relationships among these parts. For this reason, we turn next to a very brief description of the basic ingredients in any performance assessment, depicting the component parts and the order in which they can be assembled. And then, we can turn to the important considerations in the development of performance exercises.

---

This chapter has been adapted and reprinted with the permission of Macmillan College Publishing Company from *Student-Centered Classroom Assessment* by Richard J. Stiggins. Copyright © 1994 by Macmillan College Publishing Company, Inc.

## An Overall Design Blueprint

We initiate the creation of performance assessments just as we initiated the development of paper-and-pencil tests: We start with a plan or blueprint. The performance assessment plan includes three components. First, the performance assessment developer must clarify the performance to be evaluated. Second, performance exercises must be prepared. And third, systems must be devised for scoring and recording results.

The immense potential of this form of assessment becomes apparent when we consider all of the design options available within this three-part structure. Let's explore those options.

### Design Part 1: Clarifying Performance

Under this heading, the user has the freedom to select from a nearly infinite range of achievement target possibilities. We can focus performance assessments on particular targets by making three specific design decisions, addressing the kind of performance to be assessed, identifying who will be assessed, and specifying performance criteria.

**Nature of Performance.** This first design decision requires that we answer these basic questions: How will successful achievement manifest itself? Where will the evidence of proficiency most easily be found?

Performance might take the form of a particular set of skills or behaviors that students must demonstrate. In this case, we watch students "in process," or while they are actually doing something, and we evaluate the quality of their performance. For example, a primary-grade teacher might observe a youngster in interaction with other students. In this instance, success manifests itself in the actions of the student.

On the other hand, we also can define performance in terms of a particular kind of product to be created. The teacher would examine the product and determine the extent to which the criteria of quality had been met. Evidence of success is found in the attributes of the product.

Some contexts permit or may even require the observation and evaluation of both skill and product. For example, you might watch a student operate a computer (skill) and evaluate the final product (the resulting program or other document) to evaluate that student's success in hitting key parts of that achievement target.

**Focus of the Assessment.** To address this design decision, we must understand that performance assessments need not focus only on individual student behaviors. They also can apply to the observation of and judgment about the performance of students functioning in a group.

In these times of cooperative learning, the evaluation of teamwork can represent a very important and useful application. Two kinds of observations are worthy of consideration. One focuses on the group interaction behaviors. The observer tracks the manner in which the group works as a whole. The group is the unit of analysis of performance, if you will. A sample performance assessment question might be: Given a problem to solve, does the group exhibit sound group problem-solving skills?

The other form of observation focuses on individual behaviors in a group context and summarizes those across individuals. For example, observers might tally and/or evaluate the nature of instances of aggressive and dangerous playground behavior. These can be very informative and useful assessments.

**Performance Criteria.** Once we have decided on the performance and performer upon which to focus, the real work begins. Attention then shifts to (a) specifying in writing all key elements of performance, and (b) defining a performance continuum for each element so as to depict in writing what performance is like when it is of very poor quality, when it is outstanding, and all key points in between. These key elements or dimensions of performance are called the performance criteria.

Performance criteria answer the questions: What are the desirable behaviors to be exhibited? What are the specific attributes of a quality product?

Clear and appropriate performance criteria are critical to sound performance assessment. When we can provide sound criteria, we are in for an easy and productive application of this methodology. Not only will we be in focus on the expected outcomes, but with clearly articulated performance criteria, both students and teachers share a common language in which to converse about those expectations. (See Section VI of this guide for greater detail about the development of performance criteria.)

### Design Part 2: Developing Exercises

In designing performance exercises, we must think of ways to cause students to perform in a manner that will reveal their level of proficiency. How can we cause them to produce and present a sample product for us to observe and evaluate, for example? In this facet of performance assessment design, we decide the nature of the exercises, the number of exercises needed, and the actual instructions to be given to performers. I will outline these design challenges briefly below and then treat them in greater depth later in the chapter.

**Nature of Exercises.** Performance assessment offers two exercise choices that, once again, reveal the rich potential of this methodology. Specifically, there are two ways to elicit performance for purposes of evaluation.

One option is to present a structured exercise in which you provide respondents with a predetermined and detailed set of instructions as to the kind of performance desired. They are completely aware of the assessment, they reflect upon and prepare for this assignment, and then they provide evidence of their proficiency. For example, they might be asked to prepare to give a certain kind of speech, perform some kind of athletic activity, or write a term paper.

But also be advised that performance assessment offers another option not available with any other form of assessment. You can observe and evaluate some kinds of performance during naturally occurring classroom events and gather useful information about "typical" student performance. For example, a primary-grade teacher interested in the social interaction skills of one student obviously would disrupt the entire assessment by instructing the student, "go interact with that group over there, so I can evaluate your ability to get along." Such an exercise would run completely counter to the very essence of the assessment. Rather, the teacher would want to stand back and watch the student's behavior unfold naturally in a group context to obtain usable information. Assessments that are based on observation and judgment allow for this possibility, while others do not.

It may have become apparent to you also that you can combine observations derived from administration of structured exercises and from the context of naturally occurring events to generate corroborating information about the same achievement target. For example, an English teacher might evaluate writing proficiency gathered in response to a required assignment and in the context of student daily writing journals done for practice. The combined information might provide insights about specific student needs.

**Content of Exercises.** A second exercise-related design component is the instructions given to performers. Instructions for structured exercises should include the kind(s) of achievement to be demonstrated, conditions under which that demonstration is to take place, and standards of quality to be applied in evaluating performance. Here is a simple example of a complete exercise:

*Achievement.* Your four-person team is to do the research required to prepare a group presentation on the dwellings and primary food sources of the native American tribe you have selected.

*Conditions.* The manner in which you carry out the background research and divide up responsibilities within the group is up to you. The focus of the evaluation will be your presentation. (Note that the process of doing the background research or preparation will not be evaluated.)

*Standards.* Your presentation will be evaluated according to the criteria we develop together in class, dealing with content (scope, organization, and accuracy) and delivery (use of learning aids, clarity, and interest value for the audience).

**Number of Exercises.** Once the nature and content of the exercise are determined, the developer must decide how many exercises are needed. This is a sampling issue. How many examples of student performance are enough? Users must decide how many exercises are needed to provide a representative sample of all the important questions they could have asked given infinite time. If we want to know if students can speak French, how many times do they have to speak for us to be reasonably certain we could predict how well they would do given one more chance? How many samples of writing must we see to be confident drawing conclusions about writing proficiency? In fact, the answers to these questions are a function of the assessment context. To answer them, we must consider several factors, including the reasons for assessment and other issues. We will review these factors later in the paper.

### Design Part 3: Scoring and Recording Results

Once performance has been clarified and exercises developed, procedures for managing results must be specified.

**Level of Detail of Results.** First, the user must select one of two kinds of scores to generate from the assessment. Option one is to evaluate performance analytically, making independent judgments about each of the performance criteria separately. In this case, performance is profiled in terms of individual ratings. Option two is called holistic scoring. In this case, one overall judgment is made about performance that combines all criteria into one evaluation. The choice is a function of the manner in which assessment results are to be used. Some uses require the high-resolution microscope of the analytical system, while others require the less precise but also less costly holistic process.

**Recording Procedures.** Second, designers must select a specific method for transforming performance criteria into usable information through a system of recording procedures. Once again, the great flexibility of performance assessment methodology comes through. Users have many choices here, too:

- Checklists of desired attributes present or absent in performance
- Various kinds of performance rating scales
- Anecdotal records, which capture written descriptions of and judgments about performance

- Mental records, which capture images and records of performance in the memory of the evaluator for later recall and use (to be used cautiously!)

**Identifying the Rater.** And finally, performance assessment users must decide who will observe and evaluate performance. In most classroom contexts, the most natural choice is the teacher. Since performance evaluators must possess a clear vision of the desired achievement and be capable of the rigorous application of the performance criteria, who could be more qualified than the teacher?

Just be advised that we have other choices. We might rely on some outside expert to come to the classroom and participate. Or we might rely on the students to conduct self-assessments or to evaluate each other's performance. The instructional potential of preparing students to apply performance criteria in a rigorous manner to their own work should be obvious.

## The Order of Performance Assessment Development

In explaining the basic three-part design framework (summarized in Figure 1), I began by specifying performance, then turned to exercises, then scoring practices. However, this is not the only possible or viable order of development for performance assessments.

For instance, we might begin developing a performance assessment by creating rich and challenging exercises. If we can present complex but authentic, real-world problems for students to solve, then we can see how they respond during pilot test administrations and devise clear criteria for evaluating the achievement of subsequent performers.

On the other hand, it is difficult to devise exercises to elicit outcomes unless and until we have specified precisely what those outcomes are. In this case, we could start by selecting the target, translate it into performance criteria, and then develop performance rating procedures. Then, given a clear vision of the desired performance, we can devise exercises calculated to elicit samples of performance to which we can then apply the criteria.

In a sense, we have a chicken-or-egg dilemma here. We can't plan to evaluate performance until we know what that performance is—but neither can we solicit performance properly until we know how we're going to evaluate it! Which comes first, the performance criteria or the exercises?

As luck and good planning would have it, we have a choice. Which we choose is a function of our level of understanding of the valued outcomes to be assessed.

FIGURE 1
PERFORMANCE ASSESSMENT DESIGN FRAMEWORK

| **Design Factor** | **Options** |
|---|---|
| 1. *Clarifying Performance* | |
| Nature of performance | Behavior to be demonstrated<br>Product to be created |
| Focus of the assessment | Individual performance<br>Group performance |
| Performance criteria | Reflect key aspects of the specific target |
| 2. *Developing Exercises* | |
| Nature of exercises | Structured assignment<br>Naturally occurring events |
| Content of exercises | Defines target, conditions, and standards |
| Number of exercises | Function of purpose, target, and available resources |
| 3. *Scoring and Recording Results* | |
| Level of detail of results | Holistic<br>Analytical |
| Recording procedures | Checklist<br>Rating<br>Anecdotal record<br>Mental records |
| Identifying the rater | Teacher<br>Outside expert<br>Student self-evaluation<br>Student peer evaluation |

### When You Know What to Look For

Those who begin the performance assessment development process with a strong background in the area to be evaluated probably possess a highly refined vision of the target and can develop performance criteria out of that vision. If you begin with that kind of firm grounding, you may be able simply to sit at your desk and spell out each of the key elements of sound performance. With little effort, you may be able to translate each key element into different levels of achievement proficiency in clear, understandable terms. If you have sufficient pedagogical knowledge in your area(s) of expertise, you can use the procedures discussed in this chapter to carry out the necessary professional reflection, spell out the performance criteria, and transform those into scoring and recording procedures. Then you will be ready to devise exercises to elicit the performance you want.

### When You Have a Sense of What to Look For

However, not everyone is ready to jump right in as just described. Sometimes we have a general sense of the nature of the performance but are less clear on the specific criteria. For example, you might want your students to "write a term paper," but not have a clear sense about the standards of quality you want to apply.

When this happens, you need a different starting place. One option is to give students a general term paper assignment and use the resulting papers—that is, actual samples of student work—as the basis for defining specific criteria. You can select a few high-quality and a few low-quality papers to compare as a means of generating clear and appropriate performance criteria. One way to do this is to sort them into three or four piles of general quality, ranging from poor to excellent, then carefully and thoughtfully analyze why the papers differ. Why do some work, while others don't? In those differences are hidden the performance criteria you seek.

The major shortcoming of starting with general exercises, of course, is that it puts students in the unenviable position of trying to perform well—write a good term paper, for example—without a clear sense of what good performance is supposed to look like. But remember, you need do this only once. From then on, you will always have well-developed criteria in hand to share with students in advance.

You can avoid this problem if you can recover copies of previous term papers. The key is to find contrasting cases, so you can compare them. They needn't come from your current students. Or, if you are assessing a demonstrated skill, perhaps you can find videotapes of past performance, or can locate students practicing and observe them. One excellent way to find the right performance criteria—your vision of the meaning of academic success—is by "student watching." You can derive the critical elements of student success from actual samples of student work. But to take advantage of this option, you first need to get them performing. That may mean starting with the development of exercises.

### When You're Uncertain About What to Look For

Other times, you may have only the vaguest sense of what you want students to know and be able to do within a particular discipline. In these instances, you can use performance assessment exercises to help identify the truly important learnings. Here's how this works.

Begin by asking yourself, What kinds of real-world challenges do I want students to be able to handle? What are some sample problems I hope students would be able to solve? Using creative brainstorming, you and your colleagues can create and collect numerous written illustrative sample exercises. When you have assembled enough such exercises to begin to zero in on what they are sampling, then step back from the array of possibilities and ask, what are the important skills that seem to cross all or many of these problems? Or, if products are to result, what do all or many of the products seem to have in common? In short, ask, what are these exercises really getting at?

In other words, we draw inferences about the underlying meaning of success by examining various examples of how that success is likely to manifest itself in real-world problems. Out of these generalizations, we can draw relevant and explicit performance criteria.

One thing I like about this strategy is the fact that the resulting performance criteria are likely to be usable for a number of similar exercises. Good criteria generalize across tasks. They are likely to represent generally important dimensions of sound performance, not just those dimensions that relate to one narrowly defined task. They capture and convey a significant, generalizable portion of the meaning of academic success.

## Developing Quality Exercises

Now let me turn to the primary intended focus of this chapter. Performance assessment exercises, like selected response test items and essay exercises, frame the challenge for the respondent and set the conditions within which that challenge is to be met. Thus, they are clear and explicit reflections of the desired outcomes. Like essay exercises, sound performance assessment exercises outline a complete problem for the respondent: achievement to be demonstrated, conditions of the demonstration, and standards of quality to be applied.

As specified earlier, we face three basic design considerations when dealing with exercises in the context of performance assessment. We must determine the following:

1. The nature of the exercise(s), whether structured exercises or naturally occurring events
2. Specific content of structured exercises, defining the tasks to be carried out by performers
3. The number of exercises needed to provide a sufficient sample of performance

Let's now delve into each in some detail.

### Nature of Exercises

The decision about whether to rely on structured exercises, naturally occurring events, or some combination of the two should be influenced by several factors related to the outcome(s) to be assessed and the environment within which the assessment is to be conducted.

**Focus of Assessment.** Structured exercises and naturally occurring events can help us get at slightly different targets. When a pending performance assessment is announced in advance and students are given instructions as to how to prepare, we intend to maximize their motivation to perform well. In fact, we often try to encourage best possible performance by attaching a grade or telling students that observers from outside the classroom (perhaps parents) will watch them perform. When we take these steps and build the assessment around structured exercises, we set our conditions up to assess students' best possible performance, under conditions of maximum motivation to do well—a very important outcome.

However, sometimes our objective is not to see the student's "best possible" performance. Rather, what we wish is "typical" performance—performance under conditions of the students' regular, everyday motivation. For example, we want students to adhere to safety rules in the woodworking shop or the science lab all the time (under conditions of typical motivation), not just when they think we are evaluating them (maximum motivation to perform well). Observation during naturally occurring classroom events can allow us to get at the latter.

From an assessment quality control point of view, we still must be clear about our purpose. And explicit performance criteria are every bit as important here. But our assessment goal is to be watching closely as students behave spontaneously in the performance setting.

**Time Available to Assess.** In addition to motivational factors, there also are practical considerations to bear in mind in deciding whether to use structured or naturally occurring events. One is time. If normal events of the classroom afford you opportunities to gather sound evidence of proficiency without setting aside special time for the presentation of structured exercises and associated observations, then take advantage of the naturally occurring instructional event. The dividend will be time saved from having to devise exercises and present and explain them.

**Natural Availability of Evidence.** Another practical matter to consider in your choice is the fact that classrooms are places just packed full of evidence of student proficiency. Think about it: teachers and students spend more time together than do the typical parent and child or husband and wife! Students and teachers live in a world of constant interaction in which both are watching, doing, talking, and learning. A teacher's greatest assessment tool is the time spent with students.

This permits the accumulation of bits of evidence—for example, corroboration of past inferences about student proficiency and/or evidence of slow, gradual growth—over extended periods of time, and makes for big samples. It offers opportunities to detect patterns, to double check, and to verify.

**Spontaneous Assessment.** Everything I have said about performance assessment up to this point has depicted it as rational, structured, and preplanned. But teachers know that assessment is sometimes spontaneous. The unexpected classroom event or the briefest of unanticipated student responses can provide the ready observer with a new glimpse into student competence. Effective teachers see things. They file those things away. They accumulate evidence of proficiency. They know their students. No other assessor of student achievement has the opportunity to see students like this over time.

But beware! These kinds of spontaneous performance assessments based on on-the-spot, sometimes unobtrusive observations of naturally occurring events are fraught with as many dangers of misassessment as any other kind of performance assessment. Even in these cases, we are never absolved from adhering to the basic principles of sound assessment: clear target, clear purpose, proper method, sound sample, and controlled interference.

You must constantly ask yourself: What did I really see? Am I drawing the right conclusion based on what I saw? How can I capture the results of this spontaneous assessment for later use (if necessary or desirable)? Anecdotal notes alone may suffice. The threats to sound assessment never leave us. So by all means, take advantage of the insights provided by classroom time spent together with your students. But as a practical classroom assessment matter, do so cautiously. Create a written record of your assessment whenever possible.

### Exercise Content

Sound structured performance assessment exercises explain the challenge and set students up to succeed if they can, by doing the following:

- Identifying the specific kind(s) of performance to be demonstrated
- Detailing the context and conditions within which proficiency is to be demonstrated
- Pointing the respondent in the direction of a good response by identifying the standards to be applied in evaluating performance

Here is another simple example:

- **Achievement.** You are to apply your knowledge of energy converted to motion and simple principles of mechanics by building a mousetrap car, where the energy used to power your car is one snap of a mousetrap.

- **Conditions.** Using materials provided in class and within the time limits of four class periods, please design and diagram your plan, construct the car itself, and prepare to explain why you included the design features you chose.

- **Standards.** Your performance will be evaluated in terms of the specific standards we set in class, including the clarity of your diagramed plan, the performance and quality of your car, and your presentation explaining its design features. If you have questions about these instructions or the standards of quality, let me know.

In this way, sound exercises frame clear and specific problems to solve.

In a comprehensive discussion of the active ingredients of sound performance assessment exercises, Baron (1991) offers us clear and thought-provoking guidance. I quote and paraphrase below at length from her work because of the richness of her advice. Baron urges that we ask important questions about the nature of the assessment:

- "If a group of curriculum experts in my field and a group of educated citizens in my community were to use my assessment tasks as an indicator of my educational values, would I be pleased with their conclusions? And would they?" (p. 308)
- "When students prepare for my assessment tasks and I structure my curriculum and pedagogy to enable them to be successful on these tasks, do I feel assured that they will be making progress toward becoming genuine or authentic readers, mathematicians, writers, historians, problem solvers, etc.?" (p. 308)
- Do my tasks clearly communicate my standards and expectations to my students?" (p. 308)
- Is performance assessment the best method to use given what I want my students to know and be able to do?
- "Are some of my tasks rich and integrative, requiring students to make connections and forge relationships among various aspects of the curriculum?" (p. 310)
- "Are my tasks structured to encourage students to access their prior knowledge and skills when solving problems?" (p. 310)
- "Do some tasks require students to work together in small groups to solve complex problems?" (p. 311)
- Do some of my tasks require that my students sustain their efforts over a period of time (perhaps even an entire term!) to succeed?
- Do some tasks offer students a degree of freedom to choose the course of action—to design and carry out the investigations—they will take to solve the problem?
- "Do my tasks require self-assessment and reflection on the part of students? (p. 312)
- "Are my tasks likely to have personal meaning and value to my students?" (p. 313)
- "Are they sufficiently challenging for the students?" (p. 313)
- "Do some of my tasks provide problems that are situated in real-world contexts and are they appropriate for the age group solving them?" (p. 314)

These guidelines define the art of developing sound performance exercises.

### The Number of Exercises—Sampling Considerations

How do we know how many exercises are needed within an assessment to give us confidence that we are drawing dependable conclusions about student proficiency? This is a particularly troubling issue in the context of performance assessment, because the amount of time required to administer, observe, and score any single exercise can be so long. Sometimes, we feel it is impossible to employ a number of exercises because of time and workload constraints.

Consider writing assessment, for example. Because writing takes so many forms and takes place in so many contexts, defining proficiency is very complex. As a result, proficiency in one writing context may not predict proficiency in another. We understand that a proper sample of writing proficiency—one that allows generalizations to the entire performance domain—must include exercises calling for various kinds of writing, such as narrative, expository, and persuasive. Still, however, we find large-scale writing assessments identifying students as writers or nonwriters on the basis of a single twenty- to sixty-minute writing sample (Bond and Roeber 1993). Why? Because that's all the assessment resources will permit!

Sampling always involves tradeoffs between quality of resulting information and the cost of collecting it. Few have the resources needed to gather the perfect sample of student performance. We all compromise. The good news is that teachers must compromise less than the large-scale assessor primarily because they have more time with students. This is precisely why I feel that the great strength and future of performance assessment lies in the classroom, not in large-scale standardized testing.

In the classroom, it is often helpful to define sampling as the purposeful collection of a number of bits of information about student achievement gathered over time. When gathered and summarized carefully, these bits of insight can form a representative sample of performance that can lead to confident conclusions about student achievement.

Unfortunately, there are no hard and fast rules to follow in determining how many exercises are needed to yield dependable conclusions. That means we must once again speak of the art of classroom assessment. I will share a sample decision rule with you now that depicts the artistic judgment in this case, and then I will review specific factors you can consider when exercising your judgment.

The sampling decision rule is this: You know you have presented enough exercises and gathered enough instances of student performance when you can predict with a high degree of confidence how well the student would do on the next one. Part of performance assessment sampling is science and part of it is art. Dealing first with the science, the more systematic part, one challenge is to gather samples of student performance under all or most of the circumstances in which they will be expected to perform over the long haul. Let me illustrate from life.

Let's say we want to assess for the purpose of certifying the competence of commercial airline pilots. One specific skill we want them to demonstrate, among others, is the ability to land the plane safely. So we take candidates up on a bright, sunny, calm day and ask them to land the plane—clearly an authentic performance assessment. Let's say all pilots do an excellent job of landing. Are you ready to certify them?

If your answer is yes, I don't want you screening the pilots hired by the airlines on which I fly. Our assessment only reflected one narrow set of circumstances within which we expect our pilots to be competent. What if it's night, not bright, clear daylight? A strange airport? Windy? Raining? An emergency? These represent realities within which pilots must operate routinely. So the proper course of action in sampling performance for certification purposes is to analyze relevant variables and put them together in various combinations to see how each candidate performs. At some point, the array of samples of landing proficiency (gathered under various conditions) combine to lead us to a conclusion that the skill of landing safely has or has not been mastered.

This example frames your performance assessment sampling challenge, too. How many "landings" must you see under what kinds of conditions to feel confident your students can perform according to your standards? The science of such sampling is to have thought through the important conditions within which performance is to be sampled. The art is to use your resources creatively to gather enough different instances under varying conditions to bring you and the student to a confident conclusion about proficiency.

In this context, I'm sure you can understand why you must consider the seriousness of the decision to be made in planning your sample. Some decisions bear greater weight than others. These demand assessments that sample both more deeply and more broadly to give you confidence in the decision that results—such as certifying a student as competent for purposes of high school graduation, for example. On the other hand, some decisions leave more room to err. They allow you to reconsider the decision later, if necessary, at no cost to the student—for example, assessing a student's ability to craft a complete sentence during a unit of instruction on sentence construction. When the target is narrow and the time frame brief, we need sample fewer instances of performance.

Figure 2 identifies four factors to take into account in making sampling decisions in any particular performance assessment context. Even within the guidelines these provide, however, the artistic sampling decision rule is this: You know how confident you are. If you are quite certain you have enough evidence, draw your conclusion and act upon it.

FIGURE 2

## CONSIDERATIONS IN PERFORMANCE ASSESSMENT SAMPLING

- *The reason(s) for assessment.* The more critical the decision, the more sure you must be and the more information you should gather; a simple daily instructional decision that can be reversed tomorrow if necessary requires less confidence and therefore a smaller sample of performance than a high school graduation decision.
- *The scope of the target.* The broader the scope the more different instances of performance we must sample.
- *The amount of information provided by the response to one exercise.* Exercises can be written to produce very large samples of work, providing a great deal of information about proficiency; when we use these, we may need to use fewer exercises.
- *The resources available for observing and evaluating.* Put simply, the bigger your labor force, the more assessment you can conduct per unit of time. This may be something you can really take advantage of. Always remain aware of all the shoulders over which you can spread the performance assessment workload: the principal, teacher aides, colleagues, parents, outside experts, students . . .

But your professional challenge is to follow the rules of sound assessment and gather enough information to minimize the chance that you are wrong. The conservative position to take in this case is to err in the direction of oversampling to raise your level of confidence.

If you feel uncertain about the conclusion you might draw regarding the achievement of a particular student, you have no choice but to gather more information. To do otherwise is to place the well-being of that student in jeopardy.

## Some Concluding Thoughts About Exercise Development

When performance assessments are used by teachers and students in the classroom, they can serve two equally important purposes. Clearly, they can be combined with performance criteria to serve as a source of information about student achievement that can feed into a wide variety of important instructional decisions. For instance, they can help us to diagnose student needs, determine the effectiveness of instructional treatments or gather data for grading purposes.

In addition, however, they also can serve as teaching tools. To the extent that we can bring students into the performance assessment process as partners in evaluating their own and each other's achievement, we can help them to focus on the critical attributes of quality performance and as a result become better performers.

But it isn't just performance criteria that they can assist in devising and applying. They also can play a role in devising performance exercises. When they do contribute to this facet of the performance assessment development process, again, we turn an assessment event into a teaching opportunity. For example, students asked to devise performance exercises of their reasoning proficiency must come to understand the specific kinds of reasoning valued in that classroom. Otherwise they will not be capable of devising exercises that elicit those kinds of reasoning. Students striving to devise exercises that elicit particular kinds of skills or products must become masters of those targets to understand how to pose exercises that require their demonstration. In other words, the challenge of developing sound performance assessment exercises can help focus student attention on important dimensions of quality performance. Besides, when students suggest exercises, they are likely to offer suggestions of activities they are interested in pursuing. The motivational implications of that should be obvious.

Further, while the above discussion suggests that performance assessment users can be flexible in the order of development of the parts of their assessments, I am a proponent of devising all assessments around clear visions of our valued achievement targets. For this reason, I urge performance assessment users to begin the assessment development process by becoming serious students of the reasoning, skill, or product outcomes that are their instructional and assessment responsibilities. Be advised that the nature and scope of those targets cannot be and should not be merely a function of local opinion. Rather, those visions must be couched in the best current thinking of scholars and practitioners with the discipline. We can only maintain contact with that thinking by remaining in touch with the professional literature of the disciplines of our choice. In other words, we can devise performance exercises that sample student achievement appropriately when we know and understand (a) the range of applications our students might need to master within the discipline that is the focus of our assessments, and (b) what it means to perform well when making those applications.

### REFERENCES

Baron, J. B. (1991). "Strategies for the Development of Effective Performance Exercises." *Applied Measurement in Education* 4, 4: 305-318.

Bond, L., and E. Roeber. (1993). *State Student Assessment Program Database, 1992-1993*. Oakbrook, IL: North Central Regional Educational Laboratory and Washington, DC: Council of Chief State School Officers.

# V-9

# Validating and Anchoring Classroom Performance Tasks

STEPHEN K. HESS
FREDERICK COUNTY PUBLIC SCHOOLS, FREDERICK, MARYLAND

The transition to performance assessment as a primary vehicle for assessing learning and achievement requires that teachers be primary agents in designing assessment tasks. Substantial and continuous staff development is needed to ensure that they understand and use these new forms of assessment in their classrooms.

If instructional adjustment is to remain the primary reason for assessing student knowledge and skill, teachers must be satisfied that the assessments provide useful information about the specified outcomes for each task. At the design stage, certain tasks will prove to be valid instructional activities but poor assessment devices. Similarly, the utility of validated tasks is questionable if different teachers cannot agree on the relative proficiency of a given student response. Teachers, parents, and students must be assured that different teachers and schools share a common vision of the expected standards for student performance.

In Maryland's Frederick County public schools, teachers actively participate in all aspects of assessment development, from the original step of determining the essential curriculum through the subsequent stages of task design, validation, and anchoring. Effective performance assessments meet two important criteria: they assess the degree to which students have attained important learning outcomes, and they allow for consistent scoring across different classrooms and schools. The "reality check" provided by this emphasis on practitioner design has increased assessment quality and teacher ownership for the resulting assessments in Frederick County.

## Validation

Performance tasks provide evidence of how well students have attained identified learning outcomes. The first step in the validation process occurs at the design phase. Task writers attempt to "back out" of the task examples of the various kinds of responses expected from students, assuring themselves that evidence of the specified outcomes will be elicited.

More formal validation requires piloting the task to obtain actual student responses, as follows:

**Step One**—Following summer workshops, tasks are piloted with students in selected classrooms. Due to the focus on formative assessment, teachers administer the tasks in the context of their ongoing instructional program, avoiding isolated use intended to "just get it done." Concurrent with piloting the tasks, teachers should complete the Task Analysis Sheet (Figure 1) for later use by the validation team.

**Step Two**—Teachers piloting the tasks select two to four examples for each score point of the rubric and affix these examples to the Task Analysis Sheet for each task.

**Step Three**—Teachers bring the student responses and analysis sheets to a review meeting under the direction of the respective curriculum specialist or supervisor. Each participant independently reads all student responses and reviews the analysis sheets prior to meeting in subgroups by selected tasks.

**Step Four**—Breaking into subgroups for a set of tasks, each subgroup member completes the Task Review Sheet (Figure 2), independently evaluating the degree of compliance with performance task characteristics. Using the four-corner technique, each subgroup member places an *R* or *N* in a corner of a Task Review Sheet, reflecting their independent conclusion as to whether they "Recommend" (*R*) or do "Not recommend" (*N*) the task be used, folding down the corner prior to review by the next subgroup member.

**Step Five**—Following independent review by each respective subgroup member, the corners of the Anchor Summary are unfolded to reveal the degree of agreement. The intent is to reach consensus, rather than to simply add up votes, by using discussion techniques and justification of independent conclusions. In some cases, recommended tasks may require minor revisions that do not substantively change the essential task to be performed by students. In such cases, the suggested changes should be included on the Summary and presented as part of the subgroup consensus about the task.

Following completion of this process, the draft consensus is presented to the larger group of participants, with the opportunity for reaction and further consideration if the subgroup's consensus appears unclear or unjustified to the total participant group.

The intent of the task validation process is to ensure worthy and workable assessment tasks. To this end, one purpose of task validation is to identify tasks that "blew up" when attempted by students. In some cases, the entire task must be rejected. In others, some minor but important revisions may fix the problem. In either case, the resulting new or revised task must be piloted through the original process.

In cases where assessment tasks have proven worthy and workable, student responses must be anchored to the scoring rubrics.

FIGURE 1

### MARYLAND ASSESSMENT CONSORTIUM TASK ANALYSIS

Name of Task: ____________________
Subject Area: ____________________
Teacher: ____________________
Date: ____________________

Positive Attributes of task: ____________________

Problems you or the students experienced with the task:

Changes you would make the next time you used the task: ____________________

**Please attach "best examples" of student responses at each score point.**

Copyright 1991 Maryland Assessment Consortium

## Anchoring

Applying standards and criteria addressed in scoring rubrics to actual examples of student work is known as anchoring. The purpose of anchoring is to provide a set of clear examples of student responses that illustrate each score point of a rubric. Having tangible examples of student work considerably enhances the reliability of subsequent scoring by classroom teachers. In addition, anchored performances can serve as models for teachers and students, because the annotations provide instructionally relevant information. For formative assessment purposes, anchoring can be accomplished at either the task or the rubric level. (Anchoring of higher-stakes summative measures requires additional considerations not addressed by the formative focus of classroom level assessment.)

Task-level anchoring uses separate sets of anchors for each task and requires sufficient student responses for each task anchored. Rubric-level anchoring, on the other hand, provides examples of student responses from a pool of tasks that share the same scoring rubric (for example, persuasive writing). Each type of anchoring has relative merits and

FIGURE 2
TASK REVIEW SHEET

**Task:**
**Context:**
**Grade/Levels:**

| | | |
|---|---|---|
| Unrelated to Maryland outcomes | ____ | Tied directly to Maryland outcomes |
| Primary emphasis on facts/knowledge | ____ | Requires the use of higher order thinking skills and processes |
| Response limited to paper & pencil | ____ | Multiple response modes |
| Requires "expert" for scoring | ____ | Can be scored by classroom teachers |
| Excludes certain students | ____ | Can be adapted for special populations |
| Specific to one discipline | ____ | Can be adapted for multiple disciplines |
| Focus on right/wrong responses | ____ | Focus on degree of proficiency |
| Isolated performance | ____ | Contextual performance |
| Uninteresting for students | ____ | Engaging for students |
| Defined path for solution | ____ | Opportunity for creative solutions |
| Artificial and contrived | ____ | Authentic and "real world" |
| Encourages superficial responses | ____ | Requires in-depth responses |
| Taps discrete skills | ____ | Taps integrated skills |

**Comments/Recommendations:** ____________________

Copyright 1991 Maryland Assessment Consortium

disadvantages, but the latter technique encourages teachers to generalize about student performance, possibly allowing greater transfer to self-generated tasks.

The anchoring process is as follows:

**Step One**—Similar to validation procedures, the anchoring process begins with participants' reading all student responses prior to attempts at establishing scoring consensus.

**Step Two**— Total working group review of the conceptual framework for four-point rubric scales (or other scales, as appropriate):

4—Exemplary response, beyond mere proficiency

3—Proficient, but not exemplary; considerable room for further growth and development

2—Approaching proficiency, but lacking some essential element(s)

1—Evidence of attempt; little grasp of the task

NSR—Blank paper, refusal, attempted wrong task, or other nonscorable responses

Following review and discussion of the underlying framework, present and discuss the specific rubric for the set of tasks to be anchored.

**Step Three**—Affix the Anchor Summary Sheet (Figure 3) to each student response for independent review and decision by each member of the working subgroup. Using the four-corner technique described earlier, initial focus is placed on the distinction between "proficient" and "nonproficient" responses by having each reviewer place a *P* or *N* in the corner, proceeding through discussion and subgroup consensus.

FIGURE 3
ANCHOR SUMMARY SHEET

Subject/Level ____________________
Task Title ____________________

Check box reflecting working group consensus:

☐ Clear example of __________ rated response, to be included in Volume III

☐ Example of __________ rated response, consider for inclusion in Volume III

☐ Split paper without consensus on rating, do not include in Volume III

Working group members:

____________________
____________________
____________________
____________________

Copyright 1991 Maryland Assessment Consortium

**Step Four**—Following subgroup consensus on proficiency or nonproficiency, affix the Anchor Comments Sheet (Figure 4) to each "clear example" response. Working as a group, and referencing specific observations from the scoring rubric, detail the characteristics of the student's response that make it either proficient or nonproficient. When annotations are made, it is important to relate the student's response to the explicit criteria of the rubric. These

BEST COPY AVAILABLE

annotations later guide other teachers in using and interpreting the set of anchor responses when scoring student work in their classrooms.

FIGURE 4

ANCHOR COMMENTS SHEET

Subject/Level ______________________

Task Title ______________________

Consensus Rating of Paper ______________________

Annotations

______________________

______________________

______________________

______________________

Members of working group:

______________________

______________________

______________________

______________________

Copyright 1991 Maryland Assessment Consortium

After completing this process, the subgroup's draft consensus and annotations are presented to the large group for discussion and group consensus.

**Step Five**—At the discretion of each district, the initial anchoring set of proficient/nonproficient responses can be published and used by district classroom teachers, following training and opportunities to practice scoring student work. In such cases, the next phase of anchoring can await a year of proficient/nonproficient scoring.

The second phase of anchoring employs the same procedures outlined in steps one through four, but has the additional focus on differentiating proficient responses into 3's and 4's, while differentiating nonproficient responses into 1's and 2's. As a general guideline for training purposes, districts should avoid attempts to train across all four score points of a rubric until teachers can consistently differentiate between proficient and nonproficient responses.

Examples of completed forms and reaction sheets are provided at the end of this paper.

## Conclusion

The processes and procedures outlined in this article, with some individual adaptation, have been effectively used in Frederick County, Maryland, and by members of the Maryland Assessment Consortium. Obviously, there are other methods and opportunities for improvement and refinement. Regardless of the actual plan, it is critical that each district or school attempting to design performance assessments have an agreed-upon method to validate tasks and anchor student performances. Otherwise, teachers (as well as students, parents, and the community) will have little confidence in the results of the assessments or the impact of the assessments on student learning.

# VI

# *Establishing Performance Criteria*

# VI-1

# Establishing Performance Criteria

## Introduction

Judith A. Arter
Northwest Regional Educational Laboratory

### Understanding Performance Criteria

Performance criteria are used for both large-scale and classroom-level performance assessments. In large-scale assessments—those conducted for the purpose of making decisions outside the classroom—they are used to ensure consistency of scoring. They are also used by teachers in day-to-day monitoring of student progress and behavior, and by students for self-assessment.

The papers in this section are about the criteria used to judge the quality of student performances, products, or behaviors, whether used in the classroom or for large-scale assessment. The papers:

- define criteria,
- discuss the central necessity of clearly stating criteria,
- explore keys to success—characteristics of good criteria,
- consider how to develop criteria, and
- provide many examples of criteria.

### Alignment

Before discussing these topics, it is important to place performance criteria in the context of alignment (as described in Section IV). In a performance assessment, goals for students, tasks, and performance criteria must match to be effective. Goals state what it is we want to assess; also, what we want to teach. Tasks are the activities students will perform in order to demonstrate their competence in these goal areas. Performance criteria outline what we will look for in student responses to determine their progress toward the goals. For example, a goal might be to communicate well in writing. A task might require students to write an essay of two to three pages on an assigned topic. We assess the quality of the writing using criteria for what we mean by communicating well in writing, perhaps: sound ideas, good organization, personal voice, accurate word choice, sentence fluency, and correct conventions. (See the samples at the end of this chapter for writing criteria.) Or, a goal for students might be to have them become competent mathematical problem solvers. A task might involve groups of students working together to answer the question "About how many hours of advertising does the typical American 5th grader see on television in a year?" We assess the quality of responses using criteria that define "competent problem solving," perhaps: understanding the problem, using effective problem-solving strategies, correctly manipulating mathematical symbols, and clearly communicating what was done. (See the sample at the end of this section from the **Oregon Department of Education** for an example of performance criteria for mathematical problem solving.)

There are at least two implications from this. First, the task has to invite the right performance so it can be judged using the established criteria. If, for example, an assignment did not result in student writing, we could not assess students' ability to write. This might seem obvious; however, more frequently than you might think, we don't get what we thought we would when a task is assigned. (See Gail Goldberg's paper in Section V for ample examples of this.) The second implication of matching goals, tasks, and criteria is that our criteria are the final and most complete definition of our goals. We will deal with this issue in more detail below.

## Definitions

All the papers in this section address the matter of defining terms associated with performance criteria—terms such as standards, benchmarks, rubrics, descriptors, traits, and dimensions. It is important to remember that these terms are not always used consistently. For example, one author might use the term performance criteria the same way another author uses the term rubric. Or, one author might use the term trait the same way another author uses the term dimension. One author might use the word standard to mean the ideal (as in Michael Jordan set the standard for dunking basketballs), while another uses the word to mean "how much is good enough." This is just the way the field is right now; there is no true definition of any of these words.

Everyone is talking about the same sorts of things: clarifying the essential components of student products or performances (criteria, dimensions, traits); considering how strong, middle, and weak performance looks on each component (scales, descriptors); writing down our current understanding so we have a tool for communication, discussion, and scoring (performance criteria, rubric, scoring guide); determining how high performance must be on each component to be considered acceptable or excellent (standards); and illustrating what we mean with samples of student work (benchmarks, anchors, exemplars, standards).

## The Importance of Clearly Articulated Performance Criteria

The importance of performance criteria cannot be overstated. As the papers in this section (and the **Stiggins, Arter,** and **Hess** papers in Section V) clearly show, performance criteria constitute the ultimate, final definition of the vision of success we have for students. If you think of goals as general statements of what we want to accomplish with students, and expanded definitions as providing more clarity about what we mean by these goals, performance criteria are the final step in the sequence. They define what we will look for in student products and behavior to determine if they have met our goals. For example, if "ability to communicate well in writing" is a goal we have for students, then an expanded definition would outline in more detail what we mean by "communicating well in writing." Such an expanded definition might include the fact that the student crafts a piece of writing so that a clearly delineated idea is presented to the reader in a way that supports and enhances the message. This support and enhancement can be accomplished through good organization; appropriate voice; well-chosen words; a flow to the paper that makes it easy to read; and use of accurate spelling, punctuation, grammar, and capitalization.

Performance criteria, as the final step, tells us what to look for in student writing to determine if it has communicated well. If voice is part of our criteria, we must clearly define voice and specify what to look for in order to determine if the voice enhances the message of the writing. See the sample by Vicki Spandel and Ruth Culham at the end of this section for an example of performance criteria for voice and other traits that detail what to look for in writing. A related sample is that from Annie Calkins.

The papers in this section also present the case that clearly defined performance criteria are not only essential for large-scale assessment, but for instructional purposes in the classroom. **Judith Arter** maintains that clearly defined performance criteria are essential tools for communicating to students what is expected. They are also essential if we ever hope to enable students to take control of their own learning. If students don't know the criteria for success, how will they be able to judge the quality of their own work, know what to do differently next time, or track their own progress over time? Furthermore, if teachers are fuzzy about the criteria for success, or if the criteria for success differ among teachers, students are faced with double the challenge. We owe it to our students and ourselves to be crystal clear about what it takes to be successful. This doesn't just mean 90 is an *A*. When we judge the quality of performance with respect to an essay, a piece of art, oral reading, critical thinking, being a responsible citizen, or being a lifelong learner, can we clearly articulate the basis for that judgment? Do we have a commonly held vision of what it means to be successful? This vision is embodied in our performance criteria. The paper by **William L. Lazzaro** emphasizes the instructional use of performance criteria. He says student performance improves if the criteria are known, and it improves more if students help develop them.

The point here is that developing performance criteria is not an empty exercise designed so we can provide a number representing student achievement for accountability purposes. It is really an exercise in developing the vision of what we value and expect from students; an exercise in clear thinking. Thus, developing performance criteria can be a professional development activity for teachers and a learning experience for students. Also, performance criteria don't have to be something done *to* students. Students and parents can be equal partners in defining the characteristics of success.

## Keys To Success

The authors in this section look at the development of performance criteria from various perspectives and offer guidelines they've found useful in their own work. For example, all outline characteristics of performance criteria that will maximize their quality. **Joan Herman, Pamela Aschbacher, and Lynn Winters** take the approach that different features of performance criteria are appropriate for different purposes. **Grant Wiggins** and **Arter**, on the other hand, take the position that all performance criteria, regardless of the purpose of the assessment, should be designed to be classroom learning tools.

There is much agreement in these papers. All include the following on their lists of characteristics essential for quality performance criteria:

- **Don't include the trivial and don't leave out the important**. This seems obvious, but sometimes it is hard to agree on what is important, and sometimes it is hard to tell if a feature is truly the heart of a performance.
- **Be descriptive**. As Wiggins says, "The student is left with pretty weak feedback when rubrics rely heavily on words like 'less than a 5' or 'a fairly complete performance'—not much different from getting a paper back with a letter grade." We need to have a rich description that clarifies the differences between levels of performance.
- **Illustrate criteria with actual samples of student performance**. Our goal is to clarify what we mean by strong and weak performance. We should do this in any way available. Above, we discussed being descriptive rather than quantitative. Another useful way to illustrate what we mean is with actual samples of student work. When assessing or teaching voice in writing, it is very helpful to have many examples of papers with strong and weak voice so verbal descriptions take on meaning.

The papers by Wiggins and Arter add the following recommendations:

- Use general, not task-specific, criteria. Designing a rubric for each task requires too much time and energy. Task-specific criteria are also not as useful instructionally.
- **Don't use quantitative scales**. It is never right to say that a performance is better because it has, for example, more footnotes. The quality of the footnotes is as important as the quantity.
- **Use analytical trait criteria rather than holistic criteria**. Holistic criteria give a single, overall score to the whole performance. Analytical trait scales give separate scores for different dimensions of the performance. For example, in writing, one could give separate scores for ideas, organization, voice, word choice, sentence fluency and conventions. Analytical trait systems provide more diagnostic detail.

Herman, Aschbacher, and Winters are not as adamant on these three topics. Their approach is to use different types of criteria for different purposes.

The conversation on what type of criteria to use revolves around one's view of large-scale assessment. If we want only to maximize efficiency in order to provide information for accountability, some would say that we should be as efficient as possible. Holistic, task-specific criteria are very efficient—it doesn't take as long to train scorers, and scoring is fast. If, however, we see large-scale assessment as communicating important messages to teachers and students regardless of its other purposes, we should develop the kinds of systems that will maximize classroom use, even if it means less large-scale efficiency. This means we should develop generalized, analytical trait criteria. This is an ongoing debate.

## Developing Performance Criteria

The papers in this section are a good source of information on the steps involved in developing performance criteria. **Herman, Aschbacher, and Winters**, and **Arter** present similar procedures for developing generalized performance criteria. The steps are: brainstorming a list of characteristics valued in work, looking at student work to add completeness to this list, clustering this list into major categories or dimensions of performance, writing definitions of each dimension, and developing clear descriptions of what performance looks like at all levels on each dimension. All four papers propose that we base our criteria on the examination of actual student work, performances, or behavior. It is not enough to just sit down and think about what we value in student work. We get additional insights by playing our impressions against what we see when we look at actual work.

## Examples of Performance Criteria

In addition to the papers in this section discussing issues and ideas on performance criteria, we have included samples of performance criteria. They were chosen to illustrate the positive qualities described above. The samples include performance criteria for graphing, grades 5-12 (**K. Michael Hibbard**); writing, grades 3-12 (**Vicki Spandel and Ruth Culham**); science process skills, grade 9 (**Greg Hall**); science laboratory skills, grades 9-12 (**Rodney Doran**); developmental continuums in reading and writing, grades 1-2 (**Annie Calkins**); mathematical problem solving, grades 3-12 (**Oregon Department of Education**); history, grade 8 (**California Department of Education**); critical thinking, grades 9-12 (**Lake Oswego School District**); portfolios (**Spandel and Culham**); and group interaction (**Hall**).

Remember that these, like most other performance criteria, are still evolving and may not be complete, or might reflect different perspectives. Some may be analytical trait and generalized, but lacking in rich, descriptive detail. Others might be rich in descriptive detail, but holistic. (We have included no quantitative scales.) Therefore, use these criteria to provoke thinking. Do not take them as the final word on the vision you should have for student success. Above all else, be good consumers.

# VI-2

# Performance Criteria: The Heart of the Matter

JUDITH A. ARTER
NORTHWEST REGIONAL EDUCATIONAL LABORATORY

## What Are Performance Criteria?

What do you think of when you hear the word criteria? Standards? Benchmarks? Something valued? Criteria are the basis on which we judge. They provide reasons for our judgments. We use criteria in all aspects of our everyday life—every time we make an evaluative decision. They may not be written down or formalized, but they influence our decisions nevertheless. For example, we judge the quality of a restaurant on the basis of its food, ambience, cleanliness, price, service, and convenience. Our criteria for a hotel could include decor, comfort, cost, and amenities. In the world of school, we judge students on our criteria for what constitutes good work or behavior. In writing, for example, our criteria might include content, organization, voice, word choice, sentence fluency, and conventions. We tend to like writing that has a strong message or idea, is well organized, sincere and honest, fluent, and conventionally correct (proper grammar, spelling, punctuation, and capitalization). As teachers, we probably give higher grades to papers with these characteristics (commonly called dimensions or traits), even if we cannot explain why. Thus, performance criteria define and illustrate what to look for in student products or behaviors. They address concerns such as how to tell a good response from a poor one; how to be consistent in judging student performance across assignments, students, and time; how to define success; and how to capture student performance in a way to report to parents.

This paper discusses the importance of performance criteria, types of performance criteria, the characteristics of good performance criteria, and how to develop performance criteria. The goal is not only to help readers know how to develop and use performance criteria, but also to create the vision of performance criteria as "the heart of the matter." (A related paper is found in Section IV of this book, "Aligning Assessment with Curriculum and Instruction.")

## Why Are Performance Criteria Important?

Performance assessment is "assessment based on observation and judgment" (Stiggins 1994, p. 85). In performance assessment we observe a student behavior or product and make a judgment about it. Does the student demonstrate skill? How good is the product? Is student skill adequately progressing? When

---

This chapter is adapted from J.A. Arter, Chapter 3 of *Improving Science and Mathematics Education/A Toolkit for Professional Developers: Alternative Assessment*. Developed by the Laboratory Network Program—Alternative Assessment (Portland, Oreg.: Northwest Regional Educational Laboratory, 1994).

we make a judgment about anything, we use criteria whether we can articulate them or not. Our judgments in a performance assessment, therefore, are based on criteria. What happens if the criteria we use are inconsistent between raters or inconsistent in the same rater across time? What happens if raters cannot clearly articulate their criteria? If any of these is true, we run the danger of coming to different conclusions about students depending on who is doing the observing. How would this affect the credibility of the assessment and the ability of the raters to be objective?

In my experience, when developing performance assessments, the criteria for judging student performance are frequently given less attention than the design of activities for students to perform. In a review of portfolios designed for assessment purposes, Calfee and Perfumo (1993, p. 534) say:

> Respondents claim that an important purpose of portfolios is valid assessment of student progress and growth, yet nowhere in the packets have we found a clear account of how achievement is to be measured. None of the portfolio guide books . . . provides help in analysis, scoring, or grading. . . .

Clear criteria for performance are essential in a performance assessment if we are to be consistent in judging student work. However, performance criteria can be more than just a means for making consistent judgments about students. They can also be indispensable tools for learning in the classroom.

Consider the notion that criteria provide a means for clear communication between teachers and students about what is expected and what "counts." We as teachers make judgments about students all the time. Since these judgments are based on criteria, whether we can articulate them or not, we have only two choices: we can either make our criteria crystal clear to students, or we can make them guess.

> [Students have] been conditioned to believe that great papers "just happen," that they are a guessing game, and that one finds out what to do after it's too late. (Krest 1990, p. 33)

How often do we make our students guess at our criteria because we have difficulty articulating them, are too busy, or did not realize it was important?

Here are two criteria horror stories that happened to colleagues of mine this school year. The first story involves a 6th grader who was given the following problem to solve: "Three buses bring students to school. The first bus brings 9 students, the second bus brings 7 students and the third bus brings 5 students. How many students in all do the buses bring?" The student answered "21 kids" and the answer was marked wrong. After encouragement by my colleague, the student asked the teacher, "Why?" The reason was that the student said "kids" instead of "students." Having correct units can be a legitimate criterion, but only if you tell students in advance the basis for scoring. If this student hadn't been encouraged to ask, "Why?" what might have happened? The next time the student works a problem she might think, "I think this is right, but I thought I was right last time, too. Maybe I just don't understand how to do these." Not knowing what counts undermines students' confidence in their ability to judge whether they are on target and ensures that students cannot take control of their own learning.

The second story involves a high school student taking a psychology class. The student had to read one psychology article a week and write an abstract. The student consistently got *Ds* even with repeat requests to the teacher for clarification on what to do differently. This year, the student's girlfriend is taking the same class. She is turning in the same abstracts (rewritten in her own handwriting) and getting *A's*. What's the difference? The teacher didn't like the first student; she thought he was inconsiderate.

In these stories, the criteria for judging were not clear to the student, not clear to the teacher, and/or based on the wrong attributes for success. The basis on which we judge students—our criteria—are a means for communication between teachers and students. What happens if they are incomplete, unreasonable, or unclear?

The second reason performance criteria can be indispensable for learning in the classroom is that teachers are not the only ones making judgments. Students make judgments about themselves and their work without encouragement. We are now systematically asking them to self-reflect and self-assess as a way to take control of their own learning and accomplish some of the critical thinking goals we hold for them.

How often in peer review groups for writing have you felt that the feedback students give each other is superficial? Do students make comments such as "I liked it" or "It was good"? Maybe one reason for this superficiality is that students don't know what to look for and have no common vocabulary for framing their comments. I submit that if we systematically help students understand the relevant dimensions of a performance, help them recognize what works, and help them develop a vocabulary to discuss what they see, their comments will become much more sophisticated.

> I want (students) to see evaluation in its best sense—a source to inform teaching and learning. To that end we develop a vocabulary for commenting on the admirable and problematic aspects of writing . . . The more we examine samples, the richer and more

> helpful this language of evaluation becomes (Erikson 1992, p. 2).
>
> "Winning points" may be the final goal of classroom work as it is of the sports endeavor, but the grade, like the final score of the game, never taught anyone how to win again, or why they lost. For the truly successful contenders, playing the game is always about learning the game . . . however often it seems to be about scoring more wins than losses (Lucas 1992. p. 2).

The point here is that developing good performance criteria is not just an exercise in developing an assessment tool external to the instructional process. Rather, good performance criteria help teachers and students understand the targets of instruction. They help students know what is expected and what "good" looks like. Good criteria answer the questions "Why did I win?" and "How can I win again?" Performance criteria answer important questions for teachers too: "How will I know when I'm there? What kind of feedback do I give to improve student work next time?" They also provide a vocabulary for discussing work. Many teachers have reported that the most beneficial part of developing performance assessment systems in their various manifestations (portfolios and single-shot tasks) is not having the final assessment system in place, but articulating what is valued in student work and how student progress toward our goals for them (Harman 1992; Hebert 1992; Murphy and Smith 1990; *Portfolio* 1990; Sugarman 1989).

## Types of Performance Criteria

Not all performance criteria and scoring methods can be communication and self-assessment tools in the classroom. There are different types of performance criteria, and not all are designed with the instructional end-user in mind. In fact, the performance criteria designed for many assessments seem to focus more on the ease and efficiency of large-scale use than their instructional usefulness. I have reviewed more then 200 alternative assessments in science, math, writing, social studies, and reading. The following characterize the types of performance criteria I have encountered.

### Task-Specific Scoring

There are very few open-ended assessments in which performances are merely scored right or wrong. This type of scoring is usually reserved for factual types of questions requiring a short answer ("Draw a figure to show 1-3/4") Frequently, however, responses are assigned point values depending on the presence or absence of specific words, calculations, phrases, or diagrams in the response. Consider Samples 1, 2, and 3. Task-specific scoring involves developing a different set of criteria for each task. You can't use criteria developed for Sample 1 to score performance on another problem-solving task. The second task will likely not involve a board divided into four parts. This type of scoring can be used to assess different kinds of outcomes; it is not restricted to knowledge. Sample 1 was designed to assess math problem solving, Sample 2 was designed to assess

SAMPLE 1

**Task:** Mr. Ramirez helped four students make key holders. He gave Rhonda, Sam, Tony, and Uta a board and told them to share it equally. First, Rhonda measured and cut one-fourth of the board. Next, Sam measured and cut one-third of the remaining board. Finally, Tony measured and cut one-half of the remaining board. Uta used the piece that was left. Did the four students share the board equally? Draw a picture and explain your answer.

**Performance Criteria:** A "4" response contains both a picture and an explanation that indicate a clear understanding of the pattern; a picture showing a whole divided into four equal parts; or an explanation that enhances the picture by comparing the size of each piece using either sentences, computations, or a combination of both . . .

A "1" response makes some attempt at a picture or an explanation; is unclear.

*Riverside (1991)*

SAMPLE 2

**Task:** Examine the items in the garbage can. Sort the items into the following categories: paper, metal, food waste, plastic, yard waste, glass, other. Count and record the number of items in each category. List the categories in order from least number of items to greatest number of items. Complete a graph of results. (The purpose of the exercise is to see whether students can classify different types of solid waste. Students are given the same seven pieces of trash to sort.)

**Performance Criteria:** The response is scored for four different things. "*Sorting*: 7 points for all correct. Subtract 1 point for each incorrect response. *Sequencing*: 1 point for ordering items correctly; 0 points for incorrect order. *Graph*: 2 points for labeling and ordering categories on horizontal axis (subtract 1 point for incorrect order); 7 points for correct graphic representation of each category (subtract 1 point for each incorrect representation). *Solution*: 4 points for a strong, workable solution."

*Lake County ESC (1991)*

BEST COPY AVAILABLE

application of knowledge, and Sample 3 was designed to assess science process skills.

## SAMPLE 3

**Task**: Chris wants to find out which of two spot removers is better. First, he tries spot remover "A" on T-shirts that have fruit stains and chocolate stains. Next, he tries spot remover "B" on jeans that have grass stains and rust stains. Then he compares the results. What did Chris do wrong that will make it hard for him to decide which spot remover is better?

**Performance Criteria**: To get a "4" the student must recognize "kinds of material," "stains," and "spot removers" as three variables to consider; and describe control of material and stain, while varying only the spot remover.

*Kentucky Department of Education (1991-92)*

Could you tell what the samples were trying to assess before I told you? When using task-specific scoring, it is sometimes difficult to tell from looking at the task and criteria what the assessment is trying to measure. That's because assigning points for specific features is an indirect measure of performance. One infers ability from the number of points received. Thus, if students get a "4" in the key chain example (Sample 1), this is taken as an indicator of ability to problem solve. Or, if students score high on the solid waste problem (Sample 2), this is taken as evidence that students understand and can apply the concepts related to solid waste. If they get a "4" on spot removers (Sample 3), we assume they understand experimental design. This is the same logic as on a multiple-choice test. We assume that a high score indicates the student mastered whatever knowledge or skill the test is supposed to measure.

In task-specific scoring, the developers of the assessment have already made all the judgments for the rater; they have decided what constitutes good problem solving on a particular problem and have written down the rules of what, specifically, to look for. The thinking has already been done for the rater. This type of assigning points for responses is efficient and reliable, and is frequently used in large-scale assessment. You can quickly train raters to use this method. Because they are looking for the presence or absence of specifics, rating is fast and rater agreement high.

### Generalized, Judgmental Scoring

Generalized, judgmental scoring requires raters to make a direct, qualitative judgment of the skill level or proficiency shown by a student response. This type of performance criteria is not task-specific; the same criteria are used across many problems, tasks, or exercises. For example, Sample 4 shows part of the criteria used by California for judging reading, and Sample 5 (on the next page) shows part of the draft criteria being developed by Oregon to assess math problem solving. (See examples of entire rubrics at the end of this section.)

## SAMPLE 4

**California Generalized Reading Criteria**

**Exemplary Reading Performance (6 points)**

An exemplary reading performance is insightful, discerning, and perceptive as the reader constructs and reflects on meaning in a text.

Readers achieving six points develop connections with and among texts.

These readers take risks.

Readers demonstrating a six-point performance challenge the text . . .

**Minimal Reading Performance (1 point)**

In a minimal reading performance, the reader appears to understand and respond only to an individual word, title, or phrase but not to respond in ways that demonstrate even a rudimentary understanding of how these words relate to text ideas.

Any connections such readers may make to their own experience will appear in the form of words or drawings that have textual associations only to an isolated word or phrase.

Level-one reader responses suggest that these students engage in reading largely as an act of decoding rather than as a process of making coherent meaning.

*California Department of Education (1993)*

In generalized, judgmental scoring, the performance criteria are meant to directly define the skill in question so that raters can directly judge student ability in that area. Contrary to the task-specific systems above, nothing is inferred from the score. Rather, the score directly describes student ability on the skill in question. Could you tell, by the way, what these performance criteria were measuring before I told you? This is the hallmark of generalized, judgmental criteria. Since they are used to enable the rater to make a direct judgment about problem solving, they have to describe problem solving. The thinking has not been done for the rater. The rater must decide directly on the quality of the problem solving.

### SAMPLE 5

**Oregon Math Problem Solving Generalized Criteria**

**Understanding the Problem:** Understanding the problem includes the ability to interpret the problem and choose appropriate information while applying a strategy for a solution.

**Mathematical Procedures:** Mathematical procedures is the ability to demonstrate appropriate use of mathematics. It is the mathematics the student chooses to use and the selection and correct application of appropriate procedures.

**Problem Solving Strategies:** Problem solving requires the use of many skills, often in certain combinations, before the problem is solved.

**Mathematical Communication:** Mathematical communication is the use of symbols and terms which attain specific, and sometimes different, meanings to common words. Effective communication is essential to learning and knowing mathematics.

*Oregon Department of Education (1991)*

## What Do Good Performance Criteria Look Like?

As in life, there are no absolutes in assessment. No matter what guidelines you set for performance criteria, you can probably think of an exception. The following represent what I believe we should do in most cases.

### Generalized Criteria

Generalized criteria requiring direct judgment by the teacher or student are more useful as instructional tools. Task-specific criteria help you see what good performance looks like on one task, but don't help in seeing what good performance looks like in general. For example, scoring the key chain problem will not necessarily help students do better on the next task requiring problem solving. They do not help you generalize "good" problem solving. They will not help students know the rules of the game, so they can "win" next time.

Generalized criteria are just that—they help us generalize what good performance looks like across occasions.

> Scales, criteria and specific questions which students apply to their own or others' writing also have a powerful effect on enhancing quality. Through using the criteria systematically, students appear to internalize them and bring them to bear in generating new material even when they do not have the criteria in front of them (Hillocks 1986).

Generalized criteria provide an overall picture of the target, not just how the target manifests itself in a single problem or situation. Consider Sample 5. These same criteria are applied to all open-ended math problems. Evaluating one problem solution will help in evaluating the next because the goal is to understand what good problem solving is in general. The goal is generalization. Of course, in order to use the system well, raters have to see many samples of problem solutions. Thus, it takes longer to train raters than using task-specific criteria; and consistency of judgments between individuals, at least at first, is lower. In the long run, we and the students will have greater understanding if we shoot for more generalized performance criteria. Which would you rather have your students do: score a bunch of performance tasks where each task uses different criteria, or use a more generalized evaluation procedure that tries to analyze what makes problem solving good in general?

### Covering the "Right Stuff"

Good performance criteria also cover what is important, and leave out nothing of importance. I play the violin. I happen to think my style is a strength of my playing. What happens if style were left out of the criteria used to judge my playing because the judges felt it was too difficult to score reliably? What message does this send me? If teachers were held accountable for student performance, how might this affect instruction? Don't leave things, such as style in music, voice in writing, critical thinking or problem solving, out of the assessment just because they are hard to define. These are the areas *most* in need of clarification. Developing performance criteria gives us the context for direct assault on things difficult to define.

### Have Good Support Materials

Let's assume we decided that the following traits cover the important dimensions of writing: ideas, organization, voice, word choice, sentence fluency, and conventions. This doesn't help very much if we don't understand the meaning of voice. It is a common problem with performance assessment materials, especially those designed for classroom use. They give you a list of things to look for but don't help you in knowing what each looks like when you see it.

In order to be really useful, performance criteria should come with clear definitions and sample student performances to illustrate what is meant. For example, we can illustrate the trait of voice in many different ways (the more ways, the better). We can use a written definition:

> Voice is the writer coming through the words, the sense that a real person is speak-

BEST COPY AVAILABLE

## SAMPLE 6

**Sample Student Papers Illustrating Voice in Writing**

Heather hi,

What have you been up to lately. Me nothing. You have really influenced my life since I met you.

You have been my best friend since the 3rd grade you will always be. You have always been there when I needed you.

You are always there when I was feeling down and you always lifted my spirits. When are you moving back to Oregon? Well I better go. Write back soon:

Your best friend
Rochelle

*The paper "Heather hi" is very flat. The writer does not seem to be using the writing to show what they think and feel. The writer does not seem personally involved with the writing. It is generic; it could have been written by anyone. The writer does not seem to be involved in the topic. The reader is bored. The reader gets the feeling that this paper was only written because it was required. The writer really doesn't seem to have anything to say and just wants to get it over with as quickly as possible. This paper does not exhibit much voice.*

I remember spending an afternoon with "Granna" (my grandmother) before she got married. We went shopping and out for lunch. The day was hot, so we walked to an ice cream parlor and bought double scoops! When you're nine, this is very impressing. We took our double scoops to a park across the street and talked.

As I was explaining the importance of Smurfs to my grandmother, ice cream began to dribble down my arm. I jerked the ice cream to my mouth and gave a big lick, getting it all over my nose and chin. I flicked my tongue at my wrist before Granna handed me a napkin. I cleaned myself up and gave her a warm hug. "I love you." I murmured.

Now Granna is married to Roger Jamison. Roger's a nice enough guy, I guess. Very different from the gruff man I remember my grandfather to be. I love Roger, and I'm glad he's with my grandmother, making her happy. We're all very close. But sometimes, just once in a while, I wish it was just Granna and me, in the park, with ice cream on my face.

*The paper "Granna" has personality. The writer seems caught up in the topic, and enables the reader to feel what they feel. The writing is lively, personal, expressive, and individual, especially in the third paragraph. The reader really gets a feeling of love, longing, and regret from the reader. Many readers get a tear in their eye when they read the final sentence. This paper exhibits fairly strong voice.*

> ing to us and cares about the message. It is the heart and soul of the writing, the magic, the wit, the feeling, the life and breath. When the writer is engaged personally with the topic, he/she imparts a personal tone and flavor to the piece that is unmistakably his/hers alone. And it is that individual something—different from the mark of all other writers—that we call voice (Spandel, Culham, and Hill 1993, p. 5).

But we also need samples of student work with explanations of how the trait of interest is expressed. Consider the 8th grade papers and the analysis of their voice shown in Sample 6.

## Build Your Own Performance Criteria

Here's how you develop general judgmental performance criteria. I will illustrate the steps using student self-reflection as an example. Remember, we are not developing these performance criteria so we can grade student self-reflections. We are doing it as an exercise in clear thinking about one of the goals we have for students.

### Step 1: Gather Samples of Student Performance

Gather samples of student performance you feel illustrate the skill or behavior in question. In our case, we are trying to assess student self-reflection skills, so we might collect things that students have written in answer to questions like these: "Select a piece of work that you feel illustrates your best effort. Why did you select this piece? What does it show about you as a ____?" (The blank would be filled by "writer," "math problem solver," "lifelong learner," or other skill you are focusing on.) Other questions that could elicit student self-reflection might be: "How have you changed this year in your ability to ________?" or "What do you currently understand about ___?" "What do you still need to understand?" When you gather samples, it is important to represent a range of skill levels.

### Step 2: Brainstorm a List of Attributes

Place the samples of student work into three piles: strong, middle, and weak. For our example, we might have three piles that illustrate levels of sophistication with respect to self-reflection. As you sort the student work, write down your reasons for placing pieces in the various stacks. For example, if you place a piece in the "sophisticated" pile, ask yourself what you are seeing in the work that makes it sophisticated. Write down all evaluative comments; they reveal criteria.

Keep reading student work and placing it into piles until you are not adding anything new to your list of attributes. Try to create as large and diverse a list as possible. Recent lists have included such comments as these: detailed, covers many things, insightful, self-revealing, provides examples, motivation, sets goals for the future, looks at more than one thing, considers content as well as process, takes risks, is accurate, discussion relates to criteria, growth supported with examples, organized well, sincere, honest, comparisons over time, shared feelings, looks at strengths and weaknesses, depth of analysis, good reasons and explanations, revealing, voice, easy to read, looks at skill improvement, ownership, personal reaction, specific, looks ahead, thorough, synthesizes ideas, readable, neat.

When doing this initial brainstorming for a real educational application, consult the professional literature so you are as up-to-date as possible on the characteristic of interest. For instance, if you are developing criteria for critical thinking, read up on critical thinking. This preparatory reading will help you articulate the important dimensions of performance.

### Step 3: Cluster

Cluster similar attributes together and begin to form traits or dimensions of self-reflection. Try to limit the number of traits (three to six is good). For example, the above attributes might be clustered into the traits of skill analysis, sincerity, goal setting, and presentation. Make sure that your traits cover everything. For example, skill analysis might cover the following attributes: detailed, many things covered, insightful, self-revealing, provides examples, looks at more than one thing, considers content as well as process, accurate, discussion relates to criteria, growth supported with examples, looks at strengths and weaknesses, depth of analysis, good reasons and explanations, revealing, specific, thorough, and synthesizes ideas. Sincerity might cover: motivation, takes risks, sincere, honest, shares feelings, voice, and ownership. *Goal setting* might cover: looks ahead and sets goals for the future. *Presentation* might include: readable, neat, well organized, and easy to read.

### Step 4: Write a Value-Neutral Definition of Each Trait

Write a definition of each trait. For example: *Sincerity is the degree of ownership, effort and honesty in the self-reflection. It is the degree to which the student makes a sincere effort to self-analyze.* This definition should be value neutral—it describes what the trait is about, not what good performance looks like. Here is an example of the definition above converted to a statement that is not value-neutral: *Sincerity is taking ownership for work, really trying hard and being very honest; the student has made a sincere effort to self-reflect.*

### Step 5: Describe Strong, Middle, and Weak Performance on Each Trait

Generate descriptions of strong, middle, and weak performance on each trait. Use adjectives and descriptive phrases. Refer to the list of comments you made when you sorted the work. Performance criteria need to describe performance so well that you can match up what you are seeing in student work to the descriptions provided.

### Step 6: Benchmarks

Find samples of student work that are good examples of strong, weak, and mid-range. Be sure to have several sets representing typical ranges of performance at different grade levels. These can be used to illustrate to students what to do.

### EXAMPLE: SKILL ANALYSIS

| Strong (5) | Middle (3) | Weak (1) |
|---|---|---|
| • Specific, clear examples<br>• Unique<br>• Thoughtful<br>• Reflective<br>• Interesting<br>• Examines process skills<br>• Purposeful<br>• Lots of examples<br>• Tells the whole story<br>• Three-dimensional | • I like it because . . .<br>• Shows, then hides<br>• Beginning of ownership<br>• Compare/contrast<br>• What I know<br>• Two-dimensional<br>• Some specifics<br>• Describes performance | • Vague<br>• Simple restatements<br>• Mechanical<br>• Focuses only on surface features<br>• Obvious<br>• Rote response/<br>• Same old<br>• No examples<br>• Purposeless<br>• I like it/ I don't like it<br>• One-dimensional/ superficial |

BEST COPY AVAILABLE

### Step 7: Make It Better

Criteria evolve with use. Try them out. You'll probably find some parts of the criteria that work fine and some that don't. Add and modify descriptions so that they better communicate. Choose better benchmark and sample papers. If you need to, revise traits. Let students help—this is a tool for learning.

## Conclusion

If we want to maximize the impact of our performance assessments and improve their cost-efficiency, we should develop them with the instructional end-user in mind, whether they are large scale or classroom based. Assessments communicate what we value. If we value the classroom use of performance assessments and their results, value students taking control over learning rather than always having someone else judge the quality of their work, and want to promote higher-order thinking rather than just monitor it, then we have to design all performance assessments to be useful classroom tools. The bottom line is that performance criteria improve instruction if we design them properly.

### REFERENCES

California Assessment Program. (1992). *A Sampler of English-Language Arts Assessment—Elementary, Preliminary Edition*. Sacramento, Calif.: California State Department of Education.

Calfee, R.C., and P. Perfumo. (April 1993). "Student Portfolios: Opportunities for a Revolution in Assessment." *Journal of Reading* 36, 7: 532-537.

Erickson, M. (Winter 1992). "Developing Student Confidence to Evaluate Writing". *Quarterly of the National Writing Project, The Center For The Study of Writing and Literacy* 14, 1: 7-9.

Harman, S. (1992). "The Basal 'Conspiracy'". In *The Whole Language Catalog Supplement on Authentic Assessment*, edited by K. Goodman, L. Bird, and Y. Goodman. Santa Rosa, Calif.: American School Publishers.

Harvard Project Zero. (1990). "The Evolution of PROPEL." *Portfolio: The Newsletter of Arts PROPEL* 1, 6.

Hebert, E. A. (May 1992). "Portfolios Invite Reflection—from Students and Staff." *Educational Leadership* 49, 8: 58-61.

Hillocks, G. (1986). *Research on Written Composition*. Urbana, Ill.: National Council of Teachers of English.

Kentucky Department of Education. (1991-92). *Kentucky Instructional Results Information System (KIRIS)*. Frankfort, Ky.: Kentucky Department of Education.

Krest, M. (February 1990). "Adapting The Portfolio to Meet Student Needs." *English Journal* 79, 2: 29-34.

Lake County Educational Service Center. (1992). *Discovering the Problem of Solid Waste: Performance Assessments*. Grayslake, Ill.: Lake County Educational Service Center.

Lucas, C. (1990). "Introduction: Writing Portfolios—Changes and Challenges." In *Portfolios in the Writing Classroom*, edited by K. Yancey. Urbana, Ill.: National Council of Teachers of English.

Murphy, S., and M. Smith. (Spring 1990). "Talking About Portfolios." *The Quarterly of the National Writing Project & The Center For The Study of Writing and Literacy*, pp. 1-3, 24-27.

Oregon Department of Education. (1991). *Mathematics Problem Solving Scoring Guide*. Salem, Oreg.: Oregon Department of Education.

*Portfolio*. (1990). "The Evolution of PROPEL?" *Portfolio: The Newsletter of ArtsPROPEL* 1: 1. Harvard Project Zero, Harvard College of Education, Cambridge, Mass.

Riverside Publishing Company. (1991). *The Riverside Curriculum Assessment System: Performance Test Exercises*. Chicago, Ill.: Riverside Publishing Company.

Spandel, V., R. Culham, B. Hill. (1993). *Oregon Statewide Writing Assessment: 1991 and 1992*. Portland, Oreg.: Oregon Department of Education.

Stiggins, R. J. (1994). *Student Centered Classroom Assessment*. Columbus, OH: Macmillan.

Sugarman, J. (May/June 1989). "Teacher Portfolios Inform Assessment." *Harvard Education Letter* 5, 3: 5-6.

# VI-3

# Empowering Students with Instructional Rubrics

William L. Lazzaro
Mount Airy, Maryland

Learners who know how they will be evaluated on a performance assessment perform better than learners who do not have this knowledge. Putting this simple principle into practice can dramatically improve student achievement.

## Putting Research Into Practice

In the last few years, rubrics, rules, and checklists have become familiar evaluation tools for scoring performance assessments. Despite their widespread use, many educators do not fully capitalize on the instructional value of rubrics and other scoring tools. In the May 1992 *Educational Leadership*, Grant Wiggins states:

> All real world performers know the target and the standards, not just their task, in advance; such knowledge guides their training and rehearsals. Students should never have to wonder "Is this right?" "Am I finished?" "How am I doing?" "Is this what you want?" (p. 29)

Wiggins presents a solid case for allowing students to know and use the scoring rubrics and rules *before* they begin a performance assessment.

In spring 1992, I put Wiggins's recommendations into practice in my 5th grade classroom at Damascus Elementary School in Montgomery County, Maryland. My colleague, Susan J. Montgomery, and I were impressed at the increased level of student achievement we observed each time we used instructional rubrics with our students. We use the term "instructional rubrics" to identify criterion-based scoring instruments which explicitly convey to the student specific, teacher expectations for a performance-assessment task. Thus, the rubric becomes an integral part of instruction, rather than an after-the-fact evaluation tool.

During the 1992-93 school year, Susan and I began extensively using scoring rubrics in our teaching. At first, we used teacher-made rubrics. We gave students many opportunities to practice evaluating their own work with the rubrics and rules. Then we asked our students to help us write rubrics for classroom assignments, projects, and book reports. By November of that year, our students were able to write their own rubrics for independent projects. Realizing our students would be taking the Maryland School Performance Assessments in May, we taught them to analyze a performance-assessment prompt and to construct a rubric for that task. The benefit of students knowing the rubric before they began a task was reinforced time and again. We were convinced

---

Copyright © 1995 William L. Lazzaro

that we had found an instructional strategy that really made a difference in classroom instruction.

We conducted several informal studies with 1st, 4th, and 5th grade students in our school. In some cases, we taught lessons to students and gave them parallel performance tasks to complete. Each task was complex and had multiple steps. First, students completed the tasks without the benefit of a rubric. The following day, students were introduced to the concept of rubrics, assigned a parallel task, and given the rubric for evaluating performance on that task. Some examples of the student work are shown in Examples 1, 2, and 3 (beginning on p. vi–3 : 4).

First grade students read similar literature selections during their last week of school. After reading each book, students answered the same questions in writing: "Who is the main character?"; "What is the setting for the story?"; "What was the problem in the story?"; and "How was the problem solved?". Students answered the questions about the first book without a rubric. (See Figure 1 for one student's response.) The rubric (Figure 2) was explained to students before they answered the questions about the second book. With the aid of the rubric, students then produced a second set of responses. (See Figure 3 for the same student's response with the rubric.)

This student's work is typical of the writing produced by 1st graders with and without a rubric. Fourth grade students read *The Great Kapok Tree* and on different days responded to similar writing prompts. Examples 4, 5, and 6 depict one student's initial response, the rubric that was presented to students, and the same student's second response. The two examples of student work show the dramatic effect a rubric can have on student performance.

In science class, 5th grade students completed a Montgomery County Public Schools prototype performance assessment called Paper Bridges. This multistep performance task leads students through a series of investigations in which they examine ways a paper bridge can be strengthened when it is constructed with creases or folds. First, I let my students complete the task. After two weeks, I students the same task, but this time we worked together as a class to find the criteria in the prompt. Together, we underlined the criteria that characterize an excellent performance. My students completed the task again using the rubric we had identified in the prompt. Examples 7, 8, and 9 present the task description, with embedded criteria underlined, and the "before" and "after" examples of one student's responses.

## Grassroots Campaign Begins

Montgomery County Public School administrators were quick to recognize the significance of our instructional use of rubrics. They provided support for our work and arranged for other county teachers to learn about this approach. We have found that teachers respond enthusiastically to instructional rubrics because they find the idea easy to implement and can see positive results with their students.

## Results Confirmed in State Performance Assessments

The results of the 1993 Maryland School Performance Assessment Program verified Wiggins's recommendations and our year-long experiences with our class of 5th graders. Damascus Elementary School students scored notably higher than both the state average and the Montgomery County Public Schools (MCPS) average (Figure 1). In a county with 126 elementary schools, only one other school achieved higher performance levels, even though the demographic characteristics of our school would place it near the middle of the county. While a number of factors lead to student achievement, we believe these significant results were at least in part due to our students' ability to analyze the perform-

FIGURE 1

1993 MARYLAND SCHOOL PERFORMANCE ASSESSMENT PROGRAM RESULTS
(PERCENTAGE OF STUDENTS RECEIVING A SATISFACTORY LEVEL)

| Test | State | MCPS | Damascus E.S. |
|---|---|---|---|
| Reading | 24.7 | 35.8 | 77.3 |
| Math | 39.5 | 59.6 | 85.1 |
| Soc. Si. | 31.3 | 47.1 | 76.6 |
| Science | 33.3 | 50.2 | 63.8 |

ance assessment prompts, find the rubric, and use the rubric to complete the task.

There are numerous benefits of using rubrics for students, teachers, and parents. Learners are empowered to achieve success when they know the rubric before they begin a task. Because they clearly understand what is expected, they can strive to meet or exceed expectations. Learners indicate that they appreciate the fairness of this evaluation process. Students develop enhanced metacognitive processes using rubrics, because the known criteria allow for self-monitoring of work. (An excellent summary of the influence of metacognitive processes on learning is found in Wang, Haertel, and Walberg 1994.)

Growing numbers of teachers recognize the value of using rubrics as instructional tools. Teachers who give students the rubric before they begin a task report significant improvements in the quality of their students' work. In addition to improved student achievement, teachers benefit from fewer misunderstandings about how students are being evaluated. Many parents express their strong support for instructional rubrics, because they see positive influences on the quality of their children's products and performances. Parents are particularly pleased with the use of rubrics for projects and reports, since they can help their children understand and meet the teacher's specific expectations for the task.

## Challenge for Performance Assessment Administration

In light of Wiggins's suggestions, and the classroom experiences of teachers using rubrics, we are convinced that students should be given the rubric for performance assessment before they begin the task. If performance assessments are to measure what students can do, it is essential that students understand the scoring criteria so they can produce their best work. The challenge for evaluators at all levels—national, state, district, and classroom—is to make the scoring criteria known to the student. The challenge to teachers is to instructionally use rubrics to help students understand and apply the criteria to improve their performance.

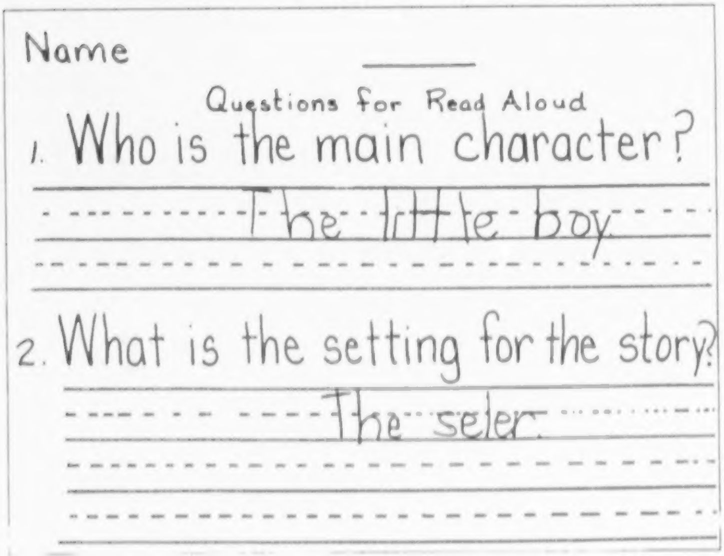
### REFERENCES

Wang, M. C., G. D. Haertel, and H. J. Walberg. (1994). "What Helps Students Learn?" *Educational Leadership* 51, 4: 74-79

Wiggins, G. (1992). "Creating Tests Worth Taking." *Educational Leadership* 49, 8: 26-33

EXAMPLE 1

## 1ST GRADE STUDENT'S WORK WITHOUT RUBRIC



Name

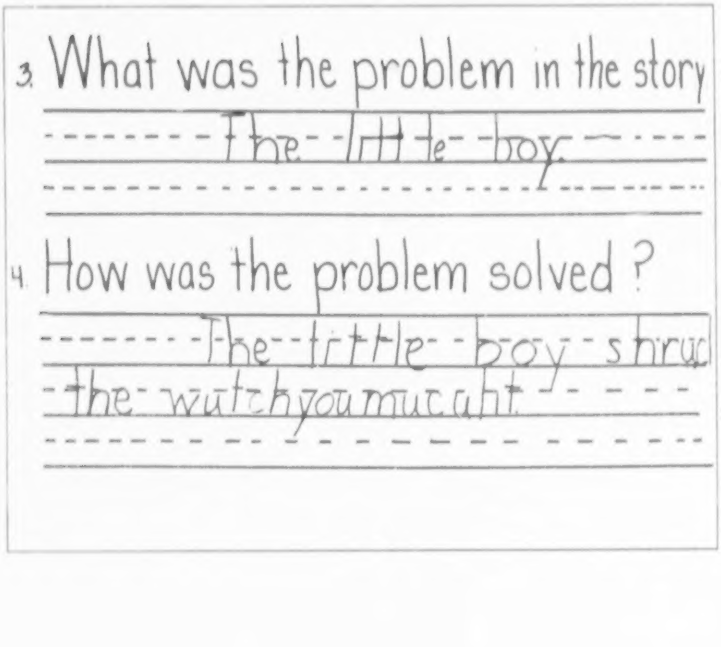
Questions for Read Aloud

1. Who is the main character?

The little boy.

2. What is the setting for the story?

The seler.



3. What was the problem in the story

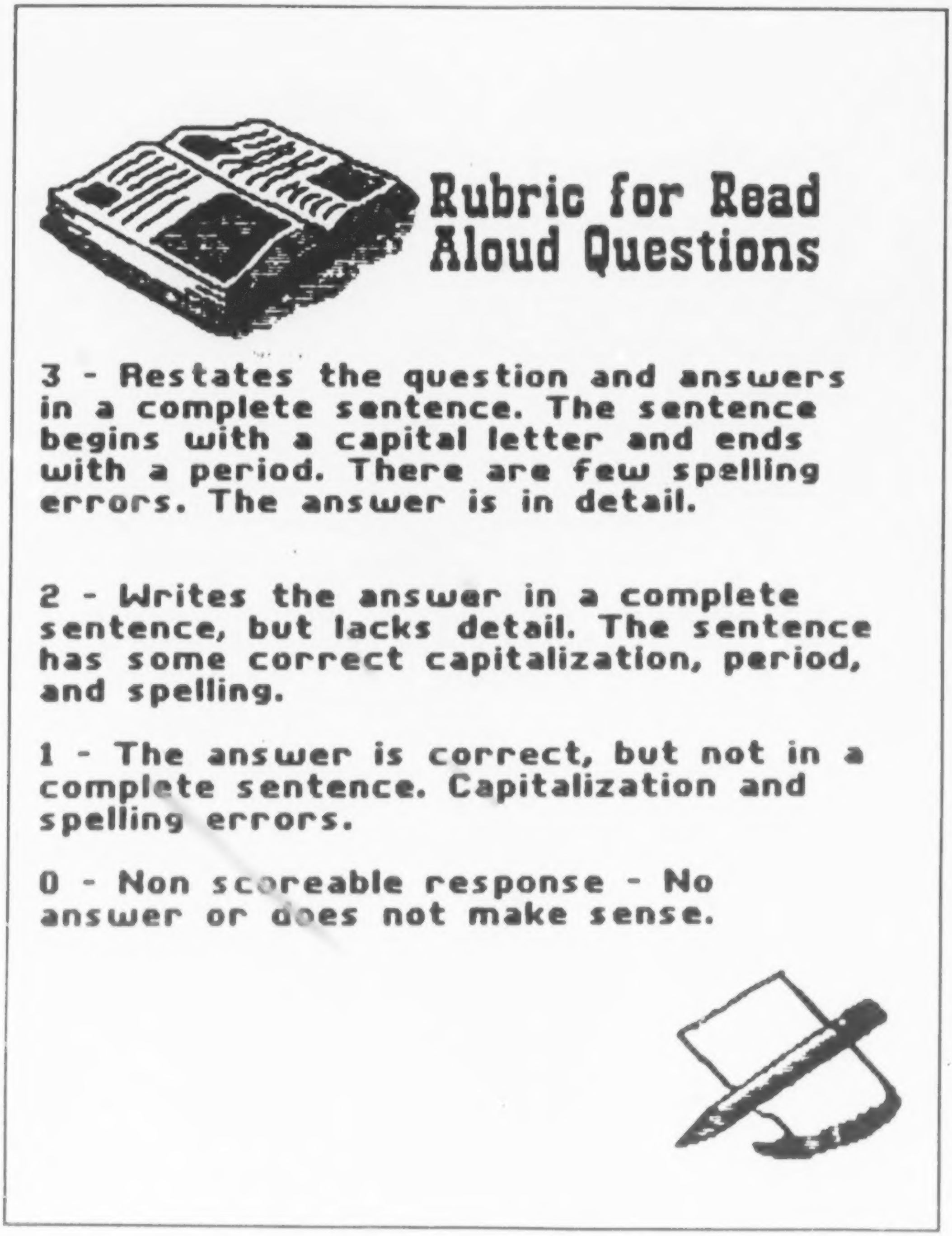
The little boy.

4. How was the problem solved?

The little boy shruc the wutchyoumucaht.

EXAMPLE 2

## RUBRIC FOR READ-ALOUD QUESTIONS



### Rubric for Read Aloud Questions

3 - Restates the question and answers in a complete sentence. The sentence begins with a capital letter and ends with a period. There are few spelling errors. The answer is in detail.

2 - Writes the answer in a complete sentence, but lacks detail. The sentence has some correct capitalization, period, and spelling.

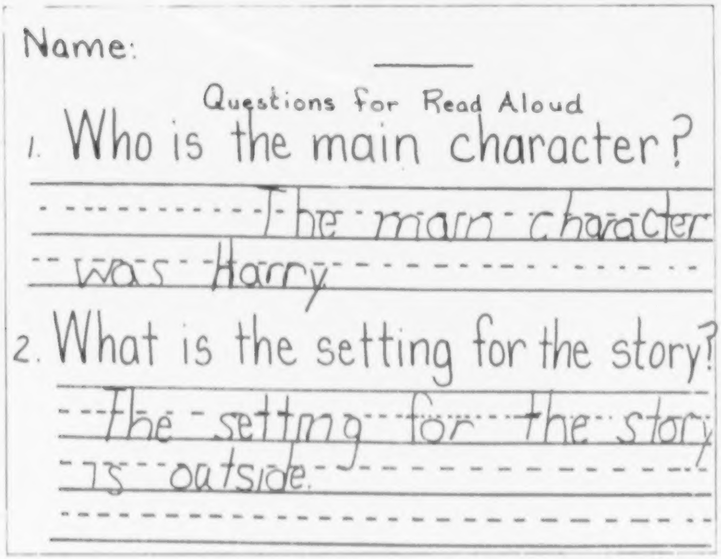
1 - The answer is correct, but not in a complete sentence. Capitalization and spelling errors.

0 - Non scoreable response - No answer or does not make sense.

BEST COPY AVAILABLE

### EXAMPLE 3

## 1ST GRADE STUDENT'S WORK WITH RUBRIC



Name:

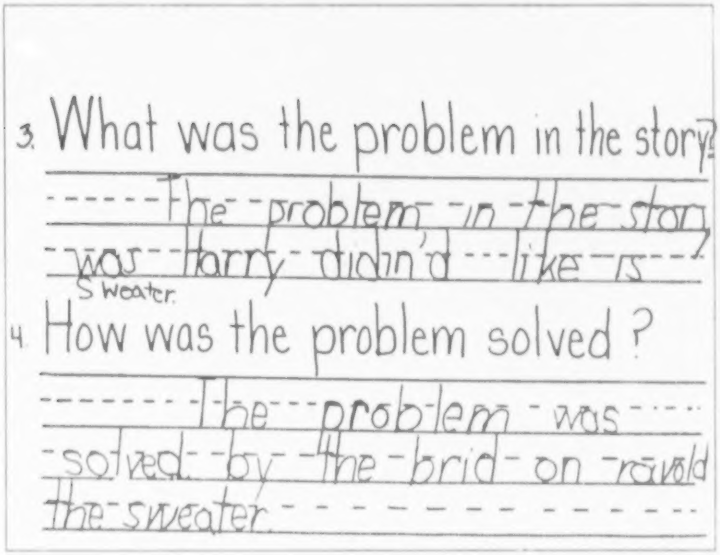
Questions for Read Aloud

1. Who is the main character?

The main character was Harry.

2. What is the setting for the story?

The setting for the story is outside.



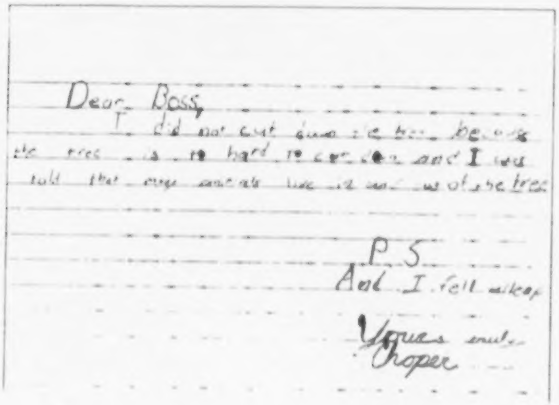
3. What was the problem in the story?

The problem in the story was Harry didn'd like is Sweater.

4. How was the problem solved?

The problem was solved by the brid on ravold the sweater.

### EXAMPLE 4
## 4TH GRADE STUDENT'S WORK WITHOUT RUBRIC



Dear Boss,
I did not cut down the tree because
the tree is to hard to cut down and I was
told that [illegible] [illegible] live in and out of the tree

P.S.
And I fell asleep

Yours truly
Choper

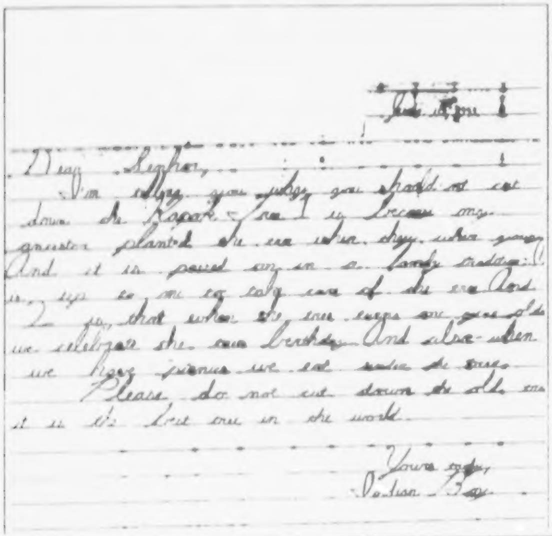
### EXAMPLE 5
## RUBRIC FOR THE GREAT KAPOK TREE

| | |
|---|---|
| 3 | • The note is written to the man to convince him not to cut down the tree.<br>• The note is written as if you were the Indian boy.<br>• More than one reason is given from the story.<br>• Correct spelling, punctuation, capitalization and grammar are used. |
| 2 | • The response is in note form.<br>• The note is written to the man to convince him not to cut down the tree.<br>• The note is written as if you were the Indian boy.<br>• Only one reason is given from the story.<br>• There are few errors in spelling, punctuation, capitalization and grammar. |
| 1 | • The response is not in note form.<br>• The response is not written as if you were the Indian boy.<br>• No reasons are given.<br>• There are many errors in spelling, punctuation, capitalization and grammar. |
| NSR | • No response given.<br>• The response does not apply to the task. |

BEST COPY AVAILABLE

### EXAMPLE 6

## 4TH GRADE STUDENT'S WORK WITH RUBRIC



[illegible]

Dear Stephen,

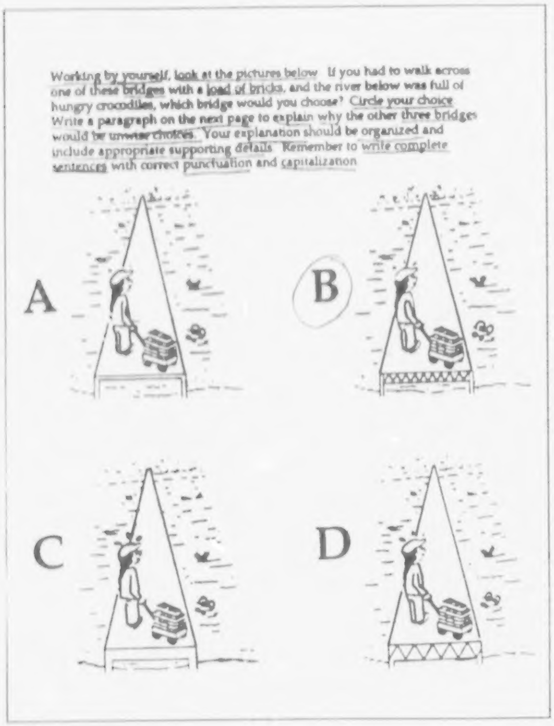
I'm telling you why you should not cut down the Kapok Tree. It is because my ancestor planted the tree when they were young. And it is passed on in a family tradition. It is up to me to take care of the tree. And I say that when the tree turns one year older we celebrate the tree birthday. And also when we have picnics we eat under the trees.

Please do not cut down the old tree it is the best one in the world.

Yours truly,

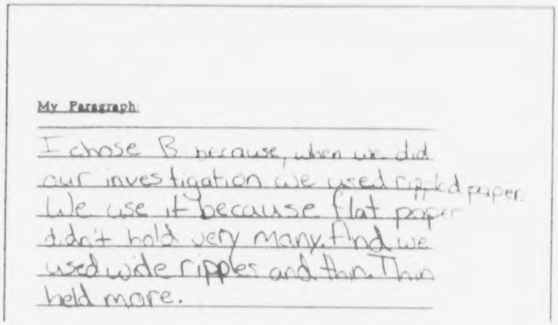
### EXAMPLE 7

## 5TH GRADE SCIENCE TASK DESCRIPTION



Working by yourself, look at the pictures below. If you had to walk across one of these bridges with a load of bricks, and the river below was full of hungry crocodiles, which bridge would you choose? Circle your choice. Write a paragraph on the next page to explain why the other three bridges would be unwise choices. Your explanation should be organized and include appropriate supporting details. Remember to write complete sentences with correct punctuation and capitalization.

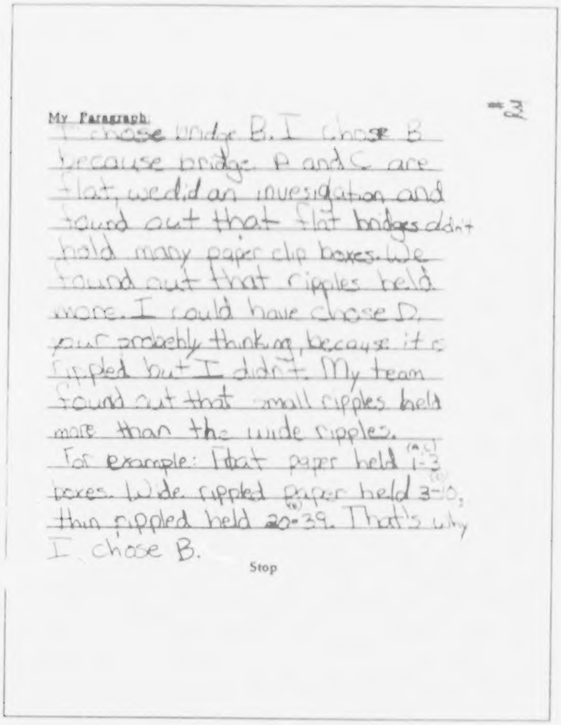
### EXAMPLE 8
## 4TH GRADE STUDENT'S WORK WITHOUT RUBRIC



My Paragraph:

I chose B because, when we did
our investigation we used rippled paper.
We use it because flat paper
didn't hold very many. And we
used wide ripples and thin. Thin
held more.

### EXAMPLE 9
## 4TH GRADE STUDENT'S WORK WITH RUBRIC



=3

My Paragraph:
I chose bridge B. I chose B
because bridge A and C are
flat, we did an invesigation and
found out that flat bridges didn't
hold many paper clip boxes. We
found out that ripples held
more. I could have chose D,
your probebly thinking, because it is
rippled but I didn't. My team
found out that small ripples held
more than the wide ripples.
For example: Flat paper held (A,C) 1-3
boxes. Wide rippled paper (B) held 3-10 (D);
thin rippled held 20-39. That's why
I chose B.

Stop

# VI-4

## Setting Criteria

JOAN HERMAN, PAMELA R. ASCHBACHER, AND LYNN WINTERS
NATIONAL CENTER FOR RESEARCH ON EVALUATION, STANDARDS AND TESTING (CRESST)

The criteria used for judging student performance lie at the heart of alternative assessment. Although we have discussed selecting and describing assessment tasks separately from developing scoring criteria, these three aspects of assessment are intimately intertwined. In the absence of criteria, assessment tasks remain just that, tasks or instructional activities. Perhaps most important, scoring criteria make public what is being judged and, in many cases, the standards for acceptable performance. Thus, criteria communicate your goals and achievement standards.

Like "alternative assessment" itself, criteria for judging student performance have been called many things, including scoring criteria, scoring guidelines, rubrics, and scoring rubrics. For our purposes, we take all these terms to mean a **description of the dimensions** for judging student performance, a **scale of values** for rating those dimensions, and, when appropriate, the **standards** for judging performance.

Let's take a common example from social studies. You assign students a group presentation accompanied by individual written reports to assess their understanding of history. Because you wish to assess three skills—oral, written, and group process skills as they relate to history—you must consider scoring criteria for each skill. Figure 1, The California History-Social Science Grade 11 Scoring Guide on the next page, is a possible set of scoring criteria for just one of these skills, a history group process assessment developed by the California Assessment Program.[1]

The group process exercise taps four **learning outcomes**: group learning, critical thinking, communication, and history knowledge. For each outcome, scoring **dimensions** are specified and levels of performance differentiated by a **scoring scale**. Finally, the scoring guide includes an **evaluation** of each performance level, labeling performance not only in terms of what was accomplished but how well, from minimal to exceptional achievement.

### Understanding the Need for Criteria

Criteria are necessary because they help you judge complex human performance in a reliable, fair, and valid manner. Scoring criteria guide your judgments and make public to students, parents, and others the basis for these judgments. Scoring a multiple-choice test does not require complicated judgment; nevertheless, human judgment is still a factor because the

---

Adapted from Chapter 5 in *A Practical Guide to Alternative Assessment* (Alexandria, Va.: ASCD, 1992), pp. 44–79.

[1] Many of the examples we use throughout this book are from state assessment programs, especially those in California. Because of their pioneering work in developing curriculum frameworks reflecting current learning and curriculum theory, certain states have already field-tested promising prototypes for alternative assessment that can be adapted for classroom use.

## Figure 1
## California Assessment Program
## History-Social Science Grade 11
## Scoring Guide: Group Performance Task

| | **Level I**<br>**Minimal Achievement** | **Level II**<br>**Rudimentary Achievement** | **Level III**<br>**Commendable Achievement** | **Level IV**<br>**Superior Achievement** | **Level V**<br>**Exceptional Achievement** |
|---|---|---|---|---|---|
| Group and Collaborative Learning<br>20 | (1-4)<br>Exclusive reliance on one spokesperson<br>Little interaction<br>Very brief conversations<br>Some students are disinterested or distracted | (5-9)<br>Strong reliance on spokespersons<br>Only one or two persons actively participate<br>Sporadic interaction<br>Conversation not entirely centered on topic | (8-12)<br>Some ability to interact<br>At least half the students confer or present ideas<br>Attentive reading of documents and listening<br>Some evidence of discussion of alternatives | (13-16)<br>Students show adeptness in interacting<br>At least 3/4 of students actively participate<br>Lively discussion centers on the task | (17-20)<br>Almost all students enthusiastically participate<br>Responsibility for task is shared<br>Students reflect awareness of others' views and opinions and include references to other opinions or alternatives in presentation and answer<br>Questions and answers illustrate forethought and preparation |
| Critical Thinking<br>30 | (1-6)<br>Demonstrates little understanding and only limited comprehension of scope of problem or issues<br>Employs only the most basic parts of information provided<br>Mixes fact and opinion in developing a viewpoint<br>States conclusion after hasty or cursory look at only one or two pieces of information<br>Does not consider consequences | (7-12)<br>Demonstrates only a very general understanding of scope of problem<br>Focuses on a single issue<br>Employs only the information provided<br>May include opinion as well as fact in developing a position<br>States conclusion after limited examination of evidence with little concern for consequences | (13-18)<br>Demonstrates a general understanding of scope of problem and more than one of the issues involved<br>Employs the main points of information from the documents and at least one general idea from personal knowledge to develop a position<br>Builds conclusion on examination of information and some consideration of consequences | (19-24)<br>Demonstrates clear understanding of scope of problem and at least two central issues<br>Uses the main points of information from the documents and personal knowledge that is relevant and consistent in developing a position<br>Builds conclusion on examination of the major evidence<br>Considers at least one alternative action and the possible consequences | (25-30)<br>Demonstrates a clear, accurate understanding of the scope of the problem and the ramifications of the issues involved<br>Employs all information from the documents and extensive personal knowledge that is factually relevant, accurate, and consistent in the development of a position<br>Bases conclusion on a thorough examination of the evidence, an exploration of reasonable alternatives, and an evaluation of consequences |
| Communication of Ideas<br>20 | (1-4)<br>Position is vague<br>Presentation is brief and includes unrelated general statements<br>Overall view of the problem is not clear<br>Statements tend to wander or ramble | (5-9)<br>Presents general and indefinite position<br>Only minimal organization in presentation<br>Uses generalities to support position<br>Emphasizes only one issue<br>Considers only one aspect of problem | (8-12)<br>Takes a definite but general position<br>Presents a somewhat organized argument<br>Uses general terms with limited evidence that may not be totally accurate<br>Deals with a limited number of issues<br>Views problem within a somewhat limited range | (13-16)<br>Takes a clear position<br>Presents an organized argument with perhaps only minor errors in the supporting evidence<br>Deals with the major issues and shows some understanding of relationships<br>Gives consideration to examination of more than one idea or aspect of the problem | (17-20)<br>Takes a strong, well-defined position<br>Presents a well-organized, persuasive argument with accurate supporting evidence<br>Deals with all significant issues and demonstrates a depth of understanding of important relationships<br>Examines the problem from several positions |
| Knowledge and Use of History<br>30 | (1-6)<br>Reiterates one or two facts without complete accuracy<br>Deals only briefly and vaguely with concepts or the issues<br>Barely indicates any previous historical knowledge<br>Relies heavily on the information provided | (7-12)<br>Provides only basic facts with only some degree of accuracy<br>Refers to information to explain at least one issue or concept in general terms<br>Limited use of previous historical knowledge without complete accuracy<br>Major reliance on the information provided | (13-18)<br>Relates only major facts to the basic issues with a fair degree of accuracy<br>Analyzes information to explain at least one issue or concept with substantive support<br>Uses general ideas from previous historical knowledge with fair degree of accuracy | (19-24)<br>Offers accurate analysis of the documents<br>Provides facts to relate to the major issues involved<br>Uses previous general historical knowledge to examine issues involved | (25-30)<br>Offers accurate analysis of the information and issues<br>Provides a variety of facts to explore major and minor issues and concepts involved<br>Extensively uses previous historical knowledge to provide an in-depth understanding of the problem and to relate it to past and possible future situations |

BEST COPY AVAILABLE

test developer phrases the questions and decides what constitutes the best answers. To the person who scores the test, a student either has or has not selected the correct answer; no judgment is needed. When we use selected-response tests, we are essentially corroborating the judgments about adequate performance built into the "answer key." Thus, all assessment, be it selected- or constructed-response, has a subjective or human judgment component.

Alternative assessments invite a wider range of possible responses. Instead of judging responses as right or wrong, alternative assessments judge the quality of, and sometimes the process of, arriving at a complex response. To make such judgments and to ensure their validity, consistency, and fairness, we need criteria or scoring guidelines. Scoring criteria must be well-conceived, explicitly defined, and consistently applied. Well-specified criteria help to ensure that everyone understands what is expected.

Well-articulated and publicly visible criteria for judging student responses are necessary and useful whether the results will be used in the classroom or to make school level or national decisions. In all assessment settings, scoring criteria must:

- Help teachers define excellence and plan how to help students achieve it.
- Communicate to students what constitutes excellence and how to evaluate their own work.
- Communicate goals and results to parents and others.
- Help teachers or other raters be accurate, unbiased, and consistent in scoring.
- Document the procedures used in making important judgments about students.

### Criteria and Instructional Planning

Scoring criteria clarify instructional goals. Along with the task description, the criteria define priority outcomes in terms of the content to be covered, the knowledge or skills to be demonstrated, and the context in which these are to occur. The complete alternative assessment specifications can guide selection and sequencing of relevant instructional activities.

### Criteria and Students

The criteria for alternative assessments are often made public and are intended to be discussed with students. Public discussions help students to internalize the standards and "rules" they need to become independent learners. Alternative assessments and their criteria can be woven into the fabric of the curriculum so that they are transparent to the student and perceived as a natural part of the learning process. Such assessment is ongoing and takes many forms—journals, conferences, peer or teacher coaching episodes, critiques of products and exhibitions, and formal evaluations of individual works or a body of work. Examples of what constitutes good work engage students in the work itself and in judgments about their work. Public discussions of quality and criteria inform students during the formative period of instruction, not simply at the end of a unit or course when it is too late to make improvements. Furthermore, discussions of criteria also help students see the perspectives of their teachers, their peers, and sometimes even the experts in the field.

### Criteria and Parent Involvement

Clearly articulated criteria also communicate to parents and others what the teachers and schools are trying to accomplish. Criteria operationalize learning goals and expectations for children. When parents know prior to grading what is expected, they can support their child's learning. For example, giving parents of kindergartners a copy of "Profile of Developmental Outcomes for Kindergarten" (Figure 2) allows them to work with their children at home on activities such as recognizing beginning letters or sight words. The road to literacy is well-marked; teachers who share the map with parents may find that more of their students reach their destinations in a timely manner.

Good criteria help both students and parents share some of the responsibility for learning. Parents and children who are familiar with the standards by which work is judged are less likely to ascribe poor performance to such external factors as not being told what was important or personality conflicts between teachers and students.

### Criteria and Consistency

When guidelines for what constitutes good work are vague or unstated, it is difficult to be consistent, fair, and accurate in judging student responses. With selected-response tests, accuracy and consistency in scoring refers to whether the test score for an individual pupil remains fairly stable from one testing occasion to another, in the absence of intervening instruction or growth. This consistency is better known as reliability. For alternative assessments, reliability includes not only the idea of the stability of an individual student's performance over time but also the stability of a rater's judgments of that performance. Specifically, a reliable assessment that depends on human judgment must meet the following requirements:

- Several judges looking at a specific task would come to the same conclusion about a student.

## Figure 2
## Profile of Developmental Outcomes for Kindergarten Literacy and Numeracy Skills

Joan C. Hillard, Superintendent, Spreckels Union School District, Spreckels, California
Elizabeth Jones, Professor, Pacific Oaks College, Pasadena, California
Jane Meade-Roberts, Director and Owner, Power of Play Preschool, Salinas, California
San Vincente School, Soledad Union School District, Soledad, California
(Jones and Meade-Roberts 1990)

| | | | | | |
|---|---|---|---|---|---|
| Oral language | Is nonverbal in school | Uses language to satisfy basic wants and needs | Often uses language in play and conversation with peers | Clearly describes real or imaginary situations using complex descriptive language | Speaks in whole sentences using a well-developed vocabulary |
| Drawing | Scribbles | Draws a face | Adds arms/legs | Adds body with arms/legs | Adds details (hair, ear, hands, etc.) |
| Writing | Scribbles and pretends to write | Uses letters or letter like signs to represent writing | Spontaneously writes own name including all letters | Spontaneously copies words | Can invent spelling of words using phonetic clues |
| Reading | Reads own name | Recognizes beginning letter of first name when written in other places | Recognizes own name, other letters and numerals | Recognizes and reads sight words, including signs, labels, key words, teacher-created word lists and/or words in books | Uses knowledge of letter sounds to sound out words |
| Attitudes toward literacy | Not yet interested in books or writing | Demonstrates focused interest in picture books | Demonstrates interest in written language (e.g., asks about or reads signs, names, words in class, labels, words in books) | Spontaneously practices writing letters and numerals | Demonstrates interest in writing correctly |
| Problem solving using classification | Randomly manipulates objects | Spontaneously orders by likenesses and differences | Recognizes or creates simple (AB) patterns using a variety of materials and/or symbols | Recognizes or creates complex (e.g., AABAAB) patterns using a variety of materials and/or symbols | Can classify by more than one attribute at a time (e.g., size and color) |
| Problem solving using numbers | Calls numerals at random | Counts by rote | Demonstrates understanding of one to one correspondence (e.g., evaluates objects accurately) | Is able to use knowledge of counting to solve real problems | Demonstrates conservation of number (e.g., understands that number of objects remains constant |

BEST COPY AVAILABLE

**Figure 2 *(continued)***

| | | | | | |
|---|---|---|---|---|---|
| Curiosity | Watches silently | Asks cautious questions | Asks questions constantly | Asks questions appropriately | Uses resources to find answers to questions (e.g., experimenting, taking risks, solving problems) |
| Creativity | Waits to be told what to do | Explores available materials | Invents a simple dramatization or projects with provided materials | Asks or looks for not already available materials to accomplish project/play idea | Works competently on notably complex, creative, imaginative, self-initiated tasks |
| Social skills with peers | Usually observes play with others | Usually plays alone or is involved in parallel play | Is developing cooperative play skills | Socially self-confident; plays effectively with other children | Has well-developed skills of leadership and cooperation in play |
| Social skills with adults/groups | Accepts situations rather than ask for adult help | Communicates with adults primarily to get help | Speaks spontaneously and freely with adults | Participates in group activities and conversation | Is sensitive to and articulate about the needs of others |
| Large motor skills | Runs | Jumps | Hops on one foot | Catches a ball with arms and chest | Can catch ball with hands only |
| Fine motor skills | Scribbles with crayon/pencil | Able to use scissors | Colors inside lines/cuts on lines | Draws/writes accurate lines | Consistently neat work |
| New learning | Chooses to observe | Prefers familiar tasks | Willing to try new tasks | Masters new tasks quickly | Masters new tasks independently |
| Social knowledge | Knows colors | Knows shapes | Knows personal information | Knows names of letters and numbers | Knows days of week, months |
| Attention span | Rapidly changing | Focuses on self-selected tasks | Focus on teacher-selected tasks | Works independently on self- and teacher-selected tasks | Can follow complex directions and maintain focused attention for long periods |

(From E. Jones and J.M. Roberts, *Profile of Developmental Outcomes for Kindergarten, Literacy and Numeracy Skills*, San Vicente School, Soledad CA)

BEST COPY AVAILABLE

- Each judge would rate the student's performance on a specific task about the same on a subsequent occasion.
- The student would perform the same task at about the same level on different occasions.
- If the task is meant to represent or generalize to some larger domain, the sample is representative of that domain.

It is easy to see how these four requirements for reliable scoring demand a mechanism for creating rater agreement and for delineating clearly the domains of particular assessment tasks. Scoring criteria must meet this demand.

### Criteria and Consequences

Specifying criteria is always important and becomes even more so when the consequences of an assessment are very serious, such as when results are used for retention, graduation, or placement in special programs. Clear guidelines for evaluating student work ensure appropriate consequences for students and the educational system as a whole. Furthermore, when alternative assessments are used for these high-stakes decisions, the scoring procedures and criteria must be legally defensible and adhere to the due process standards of a court of law.

## Specifying Criteria

Different testing purposes require different kinds of scoring criteria. Many of the examples in this book were developed for state-level assessments with such high-stakes testing purposes as comparing schools, identifying low-performing schools, and evaluating individual schools. The California Assessment Program (CAP) history group process criteria (shown in Figure 1.) are an example of the complex criteria used in high-stakes assessment. Because the criteria are used for a one-shot state assessment, the scoring guide was developed to extract the maximum amount of information possible during limited assessment time. We see that the criteria:

- List multiple learning outcomes.
- Divide each outcome into performance levels.
- Describe traits/characteristics for each level.
- Provide a numerical scale to rate the degree to which each level was attained.
- Evaluate the quality of student performance represented by the different levels using such descriptors as "minimal achievement" or "excellent achievement."

Your criteria will be less complex when your testing purposes are more focused and the decisions you wish to make about students are limited. If you are using student academic journals to monitor their progress in making connections between science lessons and their daily lives, your scoring criteria may be to count the number of unprompted statements connecting classroom learning with out-of-class experiences. The number of connections you find will tell you whether you are achieving your goals. Your assessment purpose here may be formative—to improve your instruction and to identify students who need more help or a different approach.

Perhaps your assessment purpose is more traditional—you want to evaluate student progress toward meeting your goals in mathematics problem solving. Your scoring criteria might resemble the generalized rubric for essay-type mathematics problems developed by the CAP (shown in Figure 3). The criteria provide descriptions of each level of performance in terms of what students are able to do, assign values to these levels, then apply standards at certain cut points. Students rated 1-2 are evaluated as having "inadequate" responses; students rated 3-4 receive a "satisfactory"; and students receiving 5-6 are rated "competent."

While grading is a complex issue and the scores of any one alternative assessment may or may not be used to assign grades, it is possible to find or develop criteria linked specifically to letter grades. Researchers funded by the National Science Foundation have developed a grade-linked set of criteria to assess student's procedural knowledge in a hands-on science experiment (Baxter et al. 1992). The researchers determined which methods students could use to solve the problem posed by the experiment, judged which would produce the most logical and efficient solutions, then created grade-referenced criteria to reflect their evaluations of the solutions. A summary of how their criteria is linked to grades appears in Figure 4.

Regardless of the testing purpose, the sample criteria have four common elements. Each has

- One or more traits or **dimensions** that serve as the basis for judging the student response
- **Definitions and examples** to clarify the meaning of each trait or dimension
- A **scale** of values (or a counting system) on which to rate each dimension
- **Standards** of excellence for specified performance levels accompanied by models or examples of each level.

## Considerations in Selecting Dimensions

The dimensions you use to assess student performance in a certain domain should reflect the essential qualities of good performance in that domain. Where do you find these essential qualities? The qualities or dimensions can be provided by non-educator experts, colleagues in your department, grade level teachers, district curriculum committees, re-

### Figure 3
### CAP Generalized Rubric
(California State Department of Education 1989)

#### Demonstrated Competence

**Exemplary Response . . . Rating = 6**
Gives a complete response with a clear, coherent, unambiguous, and elegant explanation; includes a clear and simplified diagram; communicates effectively to the identified audience; shows understanding of the open-ended problem's mathematical ideas and processes; identifies all the important elements of the problem; may include examples and counterexamples; presents strong supporting arguments.

**Competent Response . . . Rating = 5**
Gives a fairly complete response with reasonably clear explanations; may include an appropriate diagram; communicates effectively to the identified audience; shows understanding of the problem's mathematical ideas and processes; identifies the most important elements of the problems; presents solid supporting arguments.

#### Satisfactory Response

**Minor Flaws But Satisfactory . . . Rating = 4**
Completes the problem satisfactorily, but the explanation may be muddled; argumentation may be incomplete; diagram may be inappropriate or unclear; understands the underlying mathematical ideas; uses mathematical ideas effectively.

**Serious Flaws But Nearly Satisfactory . . . Rating = 3**
Begins the problem appropriately but may fail to complete or may omit significant parts of the problem; may fail to show full understanding of mathematical ideas and processes; may make major computational errors; may misuse or fail to use mathematical terms; response may reflect an inappropriate strategy for solving the problem.

#### Inadequate Response

**Begins, But Fails to Complete Problem . . . Rating = 2**
Explanation is not understandable; diagram may be unclear; shows no understanding of the problem situation; may make major computational errors.

**Unable to Begin Effectively . . . Rating = 1**
Words do not reflect the problem; drawings misrepresent the problem situation; copies parts of the problem but without attempting a solution; fails to indicate which information is appropriate to problem.

**No Attempt . . . Rating = 0**

search literature, and national, state, or local subject area standards committees. If you are creating criteria for your own classroom, focus your criteria on those aspects of student performance that reflect your highest priority instructional goals and represent teachable and observable aspects of performance.

One way to uncover dimensions for scoring criteria is to ask yourself the following kinds of questions:

- What are the attributes of good writing, of good scientific thinking, of good collaborative group process, of effective oral presentation? More generally, by what qualities or features will I know whether students have produced an excellent response to my assessment task?
- How does completing this task relate to my goals for students? What will they do that shows me we are working towards or achieving some of these goals?

**Figure 4**
**Linking Criteria to Grades**

| Grade | Criteria for Determining Grades |
|---|---|
| A | Student selects method.<br>Student saturates towels.<br>Student determines result so as to answer question.<br>Result logically follows from method used to saturate towel.<br>Measurements are accurate/carefully done.<br>Conclusions are correct. |
| B | Meets all requirements of an "A" but measurement is careless. |
| C | Meets all requirements of "A" but may be deficient in some areas.<br>Must attempt to control saturation by putting the same amount of water on each towel.<br>Towels not saturated (key dimension for determining a "C" or below grade). |
| D | Student fails to saturate towels or control for saturation.<br>Result is logically inconsistent with method used to saturate towels. |
| F | Student did not conduct the investigation<br>Or, equipment manipulated without purpose<br>Or, towels not wet<br>Or, conclusions based on how towels felt. |

*Criteria abridged from Baxter et al. (1992, p. 5).

- What do I expect to see if this task is done excellently, acceptably, poorly?
- Do I have samples or models of student work, from my class or other sources, that exemplify some of the criteria I might use in judging this task?
- What criteria for this or similar tasks exist in my state curriculum frameworks, my state assessment program, my district curriculum guides, my school assessment program?
- What dimensions might I adapt from work done by national curriculum councils, by other teachers?

In addition to describing your judgments about performance, the dimensions you use for your criteria need to be written so that all audiences who use them will understand them in the same way. Perhaps you are judging an interdisciplinary art project designed to reflect social studies understanding of the relationship of Native Americans to their environment. Your criteria for assigning grades or judging levels of performance should be clear to students, parents, and other teachers who depend on your judgments about content mastery, be they others at your grade level or those teaching your students next year.

Clear descriptions of performance dimensions can be achieved in several ways:

1. You could write definitions in terms of the behaviors or elements you will see when judging students. For example, instead of saying, "Acceptable performance means students show an understanding of living in harmony with the land," you could say, "Acceptable performance means that student drawings depict an environment that is almost unchanged from its original state. Few trees are cut; grassland is undisturbed except for small sustenance patches; no large waste dumps exist, and so on."
2. You could provide models or examples for each dimension. This is commonly done in direct writing assessments. Teachers are given copies of student essays exemplifying each point in the score distribution. The

essays illustrate such dimensions as, "the essay is well organized; it begins and ends effectively." From these, teachers and others can articulate precise definitions of each dimension.

3. If you are assessing informally, you could clarify your dimensions as a set of questions. For example, when you are assessing journals to see what kinds of help students need in developing fluency in writing, your criteria for deciding what to work on next could include the following questions: Which students are using some pre-writing strategies such as clustering, drawing, listing, or free-writing? Which students are keeping a log of writing ideas? Which students are having spelling problems that block the flow of ideas?

Unambiguous scale definitions usually consist of a description of the dimension to be rated, plus examples of student work illustrating acceptable responses. These models or work samples are crucial in developing a consensus about the meaning of criteria when used for rater training in formal assessments. Models also provide students with concrete examples of what acceptable or excellent work can look like. Figure 5 details one of several dimensions in a scoring rubric developed by CRESST to assess the depth of high school students' understanding of history as revealed in their essays. Note that dimensions and scale points are thoroughly operationalized: key terms, such as "concept," are defined and examples of basic points, such as statements of opinion, are provided.

In most cases, your performance dimensions, particularly for classroom assessment, will reflect your views of what constitutes excellence or expertise and will be moderated by your expectations for students at different grade levels and by your instructional goals at different points in the school year. Because your criteria help students focus on what's important instructionally, you may use different criteria at different times during the school year. For example, while you may feel that organization and mechanics are an important part of expressing discipline-based knowledge in history or science, at the beginning of the year you may particularly want to encourage fluency. Thus, your criteria at the beginning of the semester will stress the number of ideas presented, number of examples or definitions for each idea, and so on. As students become more fluent and able to substantiate their views, you can expand your criteria to include organization and mechanics. To take an example from figure skating, you may believe in the Olympic criteria of "technical merit" and "artistic expression" but at different points in your teaching you may want to differentially emphasize one or the other.

### Dimensions for Complex Tasks

It is entirely possible to create a complex assessment with multiple intended outcomes. Multiple outcomes require multiple criteria, a set for each outcome. Multidimensional criteria are unavoidable when you are doing interdisciplinary assessment or judging complex learning goals. You may either formulate separate criteria for each of these outcomes or create a multidimensional set of criteria. Connecticut's state assessment in science incorporates two approaches to assessing the same task by providing criteria for assessing group process and individual accomplishment (see Figures 6 and 7). Another perspective on student performance is provided by the subskills within the individual and group assessments. When examining group process skills, we are interested in scientific process, communication, and group collaboration. Separate criteria attend to each of these skills. The multiple dimensions on the individual scale include content and communication outcomes.

The dimensions for each scale require a lot of inference. Both teachers and students would need further descriptions of such dimensions as "draw reasonable conclusions" or "collaborate effectively" in order to use the scales. In fact, these scales are used in classrooms only after teachers have had inservice training to discuss the meaning of the dimensions, review examples, and practice using the criteria. Through classroom discussion and examples students and teachers come to a mutual understanding of the dimensions of the individual scale.

A less complex example of multidimensional criteria appears in Figure 1. The criteria assess four group performance outcomes: collaboration, critical thinking, communication, and history knowledge. The criteria include sub-criteria for deciding at which of five performance levels we should place students for each outcome. The entire set of group process criteria may be viewed as a compendium of four sets of criteria, one for collaboration, one for critical thinking, one for communication, and one for history knowledge.

## Using Rating Scales

All sample scoring criteria included in this chapter contain some type of scale, either numerical, qualitative, or both. The criteria in Figure 1, the history group process, and Figure 3, the mathematics problem, contain both numerical and qualitative rating scales. Figure 4, the hands-on science criteria, and Figures 6 and 7, the group and individual science experiment, have qualitative ratings only, such as

**Figure 5**
**CRESST Content Area Explanation**
**Essay Scoring Guidelines**
(Baker, Aschbacher, Niemi, and Sato 1992)

**CRESST Scoring Rubric Scales:**
- General Impression—Content Quality
- Number of Principles or Concepts
- Prior Knowledge: Facts and Events
- Argumentation
- Misconceptions
- Text details

**Example of Guidelines for the *Number of Principles or Concepts* Scale:**

**Number of Principles/Concepts**

This is a measure of the number of different social studies concepts or principles that the student uses with comprehension.

A *concept* is an abstract, general notion, such as "inflation." It does not refer to particular events or objects (such as one particular period of inflation), but instead represents features common to a category of events or objects. "Imperialism," for example, does not refer to any specific facts or events; it is a heading that characterizes a class of behaviors and beliefs. "Industrialization" likewise identifies a class of activities and events that share common properties. It must be clear that the student is using a term conceptually, not just as a label.

A *principle* is a rule or belief used to justify an action or judgment, as in the statement "Slavery is immoral," where "morality" serves as a justifying principle.

It should be evident that the student understands the concept and means to discuss it. The concept should not simply be mentioned within a quotation from the text with no indication that the student grasps the concept. To earn a score point, the concept or principle need not be named explicitly, such as, "Constitutionality was an important principle that influenced the debate over slavery," but the idea should be stated clearly, for example, "One problem was determining what the constitution said about slavery."

*Score point guidelines:*
0—no response
1—no concepts/principles
2—one concept/principle
3—two concepts/principles
4—three concepts/principles
5—four or more concepts/principles

Example: "One great factor that held us back from war was our economy. It was not known what would happen to our economy without the safety of Britain. Britain could defend our commerce and coasts. Also, with Britain there was a great advantage with exportation. It seemed our economy could only suffer without the aid of Britain."

Figure 6
PART II: Objectives Rating Form—Group

Student I.D. #'s
1. ____________
2. ____________
3. ____________
4. ____________
5. ____________

Title of the Task: ____________ Task # ____________

Teacher ID # ____________ Date ____________

| **The group should be able to…** | Where to Find Evidence | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Group Report (Page #) | Oral Presentation | Teacher Observation | Other (Specify) | E | G | N.I. | * |
| 1. Identify and apply physical and/or chemical properties for the purpose of identification. | | | | | ○ | ○ | ○ | |
| 2. Formulate predictions based on prior knowledge. | | | | | ○ | ○ | ○ | |
| 3. Identify information and steps needed to solve a problem. | | | | | ○ | ○ | ○ | |
| 4. Test predictions. | | | | | ○ | ○ | ○ | |
| 5. Gather data pertinent to a problem. | | | | | ○ | ○ | ○ | |
| 6. Make inferences based on pertinent data. | | | | | ○ | ○ | ○ | |
| 7. Draw reasonable conclusions and defend them rationally. | | | | | ○ | ○ | ○ | |
| 8. Communicate the strategies and outcomes of a study through written means. | | | | | ○ | ○ | ○ | |
| 9. Orally communicate the strategies and outcomes of a study. | | | | | ○ | ○ | ○ | |
| 10. Collaborate effectively. | | | | | ○ | ○ | ○ | |
| * Check if students' work is a strong and clear example of rating given. | | | | | ○ | ○ | ○ | |

(Connecticut Department of Education 1990) E = Excellent G = Good N.I. = Needs Improvement

Figure 7
PART II: Objectives Rating Form—Individual

Title of the Task ____________ Task # ____________ Student ID # ____________

Teacher ID # ____________ Date ____________

| **The group should be able to…** | Where to Find Evidence | | | | E | G | N.I. | * |
|---|---|---|---|---|---|---|---|---|
| 1. Identify and apply physical and/or chemical properties for the purpose of identification. | | | | | ○ | ○ | ○ | |
| 2. Identify information and steps needed to solve a problem. | | | | | ○ | ○ | ○ | |
| 3. Communicate the strategies of a study through written means. | | | | | ○ | ○ | ○ | |
| | | | | | ○ | ○ | ○ | |
| | | | | | ○ | ○ | ○ | |
| | | | | | ○ | ○ | ○ | |
| | | | | | ○ | ○ | ○ | |
| | | | | | ○ | ○ | ○ | |
| | | | | | ○ | ○ | ○ | |
| * Check if students' work is a strong and clear example of rating given. | | | | | | | | |

(Connecticut Department of Education 1990) E = Excellent G = Good N.I. = Needs Improvement

BEST COPY AVAILABLE

letter grades or evaluations such as "excellent" or "needs improvement."

Why scales? How do you know whether to use numerical or qualitative ratings? What about using a checklist instead of a rating scale? Whether you rate the presence or absence of a performance, as in a checklist, or use numbers or qualitative evaluations will depend on your testing purpose. There are three major types of scales: checklists, numerical ratings, and qualitative (either descriptive or evaluative) ratings. If your purpose is to **describe** what students can do, perhaps for parent conferences or to compare student performance to certain developmental standards, you may be able to use the simplest rating scale of all, the checklist. If you need more information than simply whether or not a student is engaged in specific aspects of a task, you will need a more fully developed rating scale. When you want to know the **extent** to which dimensions were observed or the **quality** of the performance, you need more elaborate scales. Rating scales, beyond the yes-no checklist format, reflect aspects of student performance other than mere accomplishment of an activity.

## Checklists

A checklist is a list of dimensions, characteristics, or behaviors that are essentially scored as "yes-no" ratings. A check indicates that either the characteristic or behavior was present or absent. Checklists often contain more dimensions to be scored than do rating scales, but those dimensions are often quite narrow and concrete.

Checklists can be useful in assessing processes, an important purpose for teachers concerned with the how as well as the what of learning. A process checklist for a hands-on experiment could resemble Figure 8, which asks the rater to note the presence of specified behaviors.

Primary school teachers find checklists useful because they must often determine how students are developing according to some theory of skills acquisition. For example, current language acquisition theory suggests that this skill cluster supports a child's ability to read:

- Ability to draw or depict an idea
- Ability to recognize sound-letter correspondence
- Ability to recognize that words stand for something
- Knowledge of left to right and up-to-down page orientation
- Ability to recall and retell favorite stories

The teacher can document acquisition of these readiness skills with a checklist. There is no need to judge how well each of these behaviors are displayed, only that they are in place. Figure 2 demonstrates a developmentally-based profile for kindergartners created by teachers of the Soledad Union School District in California, with consultation from Pacific Oaks College in Pasadena, California. This is an example of a theory-based profile. The profile development process was designed to help staff better understand constructivism, the developmental learning theory on which it is based. The behaviors identified in Figure 2 are sequenced from left to right in the order that the kindergarten staff predicted that those behaviors are acquired. This document was designed to be re-analyzed each year as teachers observe children's behaviors from a developmental point of view.

**Figure 8**
**Process Checklist**

| Procedure | Check if Observed | Comments |
|---|---|---|
| Selected approach | | |
| Correct equipment used | | |
| Measurement accurate | | |
| Sought peer help if needed | | |
| Recorded observations | | |
| Cleaned up after experiment | | |

### Numerical Scales

A numerical scale uses numbers or assigns points to a continuum of performance levels. The length of the continuum or the number of scale points can vary, three points, four points, five points, seven points—any number is possible. How many divisions or scale points should a good scale include? While there's no single answer to this question, our experience suggests that you consider these issues.

The number of points or divisions on a scale can and should vary depending on what decisions you will be making about students and whether the scale will be used in the classroom or in a formal scoring session with several raters involved in judging performance. In general, the larger the scale, the more difficult it is to clearly differentiate among the score points. Consider how quickly you can sort essays into stacks worthy of zero points, one point, or two points; essentially a decision among low, medium, and high. Why use a ten-point scale if you really only want to distinguish two or three groups of students, such as those who need additional instruction on writing a well-organized essay and those who don't?

A scale with only a few points does have some disadvantages. More scale points enable you to identify small differences between individual students and may provide more diagnostic information than a reduced scale. For example, a longer scale may be needed if you want to use one scale for all students K–12 and you also want to differentiate among students in a single grade. Also, if your scale will be used for formal assessment purposes where several readers will be rating each performance, any statistics you have to calculate, such as rater agreement, will be affected by the scale range. Using a shorter scale will result in a high percent agreement, but it will be more difficult to achieve a high correlation between raters' scores (two different ways of figuring inter-rater reliability).

It takes longer to arrive at consensus about how to assign scale points when there are more points to consider. With a five- or six-point scale, raters often refer to prior experience and assign the lowest points to off-task or truly terrible performances, the highest to stellar examples, reserve the middle for "passing," "acceptable," or model performances, then allocate those not fitting into the three anchor points to the remaining scale values. An eleven- or seventeen-point scale makes it more difficult for raters to anchor their judgments in prior experience. However, you will often see scales in multiples of five, such as ten, fifteen, or twenty point scales, which allow readers to "chunk" the points into five-point intervals. Initial rating distinctions are then really made between a five and a ten rather than a four and a seven with examples not clearly fitting into the increments receiving the intermediate points.

Another consideration related to scale size concerns multidimensional criteria. If you are rating the same performance with several criteria, each assessing a different outcome, you may want to use the same number of scale points for each outcome. Not only does this make it possible to aggregate or compare the results of several scales, but it eases the rating task. For example, using a four-point scale for coherence and a five-point scale for supporting facts could slow the rating process while raters mentally shift to different scale points. Students trying to understand their relative strengths and weaknesses can also have difficulty comparing different scales. However, if you want some outcomes to count more than others for a total score, you can use different size scales to reflect relative value or weight. A good example of this strategy appears in Figure 1, the history group process task. The scoring guide uses two different scales with one set of outcomes "weighted" up to twenty points and the other up to thirty.

### Qualitative Scales

A qualitative scale uses adjectives rather than numbers to characterize student performance. These scales are of two general sorts, descriptive and evaluative. Descriptive scales label student performance but don't necessarily make explicit the standards underlying the judgment; they use fairly neutral terms to characterize performance. Judgments about task completion, task understanding, or the appearance of certain elements in the performance are typical descriptors. Figure 9 provides three examples of descriptive scales that do not evaluate the worth of student performance.

Evaluative scales incorporate judgments of worth anchored in underlying standards of excellence. The most commonly used evaluative scales are grades (see Figure 4). Scales using descriptors of "excellence" (Figures 1, 6, and 7) or judging competence (Figure 3) are evaluative in nature. Evaluative scales require higher levels of inference to interpret than descriptive scales. The inferences are made by referring directly to the scoring criteria. The criteria themselves embed notions of excellence, competence, or acceptable outcomes.

### Numerical-Qualitative Scales

Numerical scales are often easier for people to remember, to aggregate, and to average, but are difficult to interpret in the absence of good descriptors. After all, a score of "4" on a six-point scale may connote different levels or qualities of attainment to different people. Good criteria often include both descriptive and numerical values. For example, Figure 3 displays a draft of a scale used by the California Assessment Program for judging open-ended math

**Figure 9**
**Descriptive Sales**

No evidence...Minimal evidence...Partial evidence...Complete evidence.

Task not attempted...Partial completion...Completed...Goes beyond.

Off task...Attempts to address task...Minimal attention to task...Addresses task but no elaboration...Fully elaborated and attentive to task and audience.

problems. Note that it is both numeric and descriptive. Performance is rated numerically, but each numerical score is attached to an evaluation ranging from "inadequate" to "competent."

Whether your scale values are numerical, descriptive, or both, it is important to make sure that scales help parents, students, teachers, administrators, and policymakers understand the meaning of the performance in the same way. This common understanding helps ensure reliable and fair judgments.

### The Link with Standards

Nearly all criteria, even descriptive checklists, are linked in some way to standards—the expectations for student performance. Grades or qualitative ratings reflect teacher judgment, or in the case of the hands-on science criteria in Figure 4, the consensus of the rating team. The standards underlying different scales may reflect either criterion-referenced or norm-referenced approaches to judging quality. The mathematics criteria (Figure 3) with descriptors for "inadequate response," "satisfactory response," and "demonstrated competence," reflect an absolute standard or mastery approach to standard setting. The descriptors clearly indicate good or desired performance levels, "satisfactory and above," versus poor levels, "inadequate." The levels are referenced to discipline-based standards, mathematics teachers' conceptions of adequate problem-solving strategies.

Another example is Illinois' six-point writing assessment scale, which employs an absolute scale and is designed to be used across grade levels. A score of six represents an extremely high level of writing, and few if any elementary students are expected to score above a "3." This type of scale is especially useful in measuring growth over years. The limitation of an absolute scale for multigrade/age assessment is that because elementary students all tend to score near the bottom of the scale, there is little variability in their scores so it is impossible to tell much about them individually from their scores. They all "look alike."

Other evaluative scales reflect norm-referenced approaches to standard setting. When grades or points are assigned by comparing students' relative status, such as, 'Maria's essay was better than the class average'," "Gary's video was among the best in the class," the standards are norm-referenced. Developmental checklists or scales demonstrate another common use of norm-referenced scales in alternative assessment. The sequencing of behaviors in these scales rests on what educators and others have observed over time to be typical performance at specified ages. For example, children who score "average" in reading readiness demonstrate behaviors typical for their age or grade level. "Below average" or "developmentally delayed" refers to performance typical of children in a younger age group than those being assessed.

It is possible to anchor standards in both criterion- and norm-referenced information for the same assessment. You start with a criterion-referenced scale, a scale describing performance relative to a clearly defined set of behaviors, then gather or otherwise obtain data about how a national, state, or local sample of students performed on the same measure. You can then say "Maria wrote a well-organized essay, receiving a '4' in organization; her performance was described as better than 75 percent of the students in the state." Or, on a more informal level, in your classroom, you can always describe an individual student's performance level in comparison to the rest of the class's performance: "Maria's score put her among the best in the class."

Some scales may look like absolute or criterion-referenced scales but might actually incorporate both norm- and criterion-referenced information. An age- or grade-related scale defines student performance in terms of benchmarks or expectations for a particular grade level. Benchmarks for 5th grade mathematics problem solving will differ from those for the 7th grade. What constitutes excellence in essay organization at the 8th grade will not do so at the 11th. Despite their "criterion-referenced" appearance, scales tied to an age or grade level curriculum have an underlying norm-referenced interpretation. The dimensions themselves were derived from what students were able to do at particular grades, not from absolute standards of performance across ages and grades. For practical purposes, these grade level scales are considered criterion referenced because their primary use is to decide what students can do vis-a-vis particular content and skills rather than to compare them to each other.

How can you get the best of both worlds? By determining appropriate standards according to your assessment purposes. For classroom or schoolwide assessment use, you'll probably lean toward criterion-referenced or absolute standards. For selection decisions in which there are more candidates than available space, you will probably use absolute standards for inclusion in the candidate pool, but normative standards for the final selection. For example, if you are selecting horn players for the after-school honors band, you will choose only the top 2 percent of the candidates.

We have not discussed how standards are set. How do you know where to set the acceptable level of performance? How good is competent? What is the cut point between barely satisfactory and satisfactory? High-stakes assessments, such as graduation certification, use formal standard-setting procedures. These may include using a group of judges, provided with norm- and criterion-referenced information, to determine a passing score. In district or schoolwide assessment, passing scores or labels describing poor and excellent performance are determined by consensus of those using the assessment. In the classroom, teachers set standards based on their experiences, their knowledge of what students have done in the past, their familiarity with expectations in a discipline, the current performance of students, and the purpose of the assessment.

## Considering Other Choices: Holistic or Analytic Criteria*

Based on experience with direct writing assessment, we offer two more choices in specifying criteria: holistic and analytic. Holistic criteria require raters to assign a single score based on the overall quality or to one aspect of the student's response. An analytic scale requires that raters give separate ratings to different aspects of the work. Criteria incorporating several outcomes are analytic.

### Which Is Better?

By this time you know that we're going to say "it depends on the purpose of the assessment." The pattern of results from an analytic scale provides useful feedback about the strengths and weaknesses of the individual student and the classroom instructional program. Unfortunately, because student performance on different dimensions of an analytic scale may be related in complex ways, the results may not be as clearly diagnostic as desired. Despite the fact that one of the qualities of a good analytic scale, from an efficiency and measurement perspective, is that each dimension be distinct, the subscale scores are often highly interrelated and not well differentiated. CRESST research on analytic scoring scales found high correlations among scores for overall essay and paragraph organization, and between organization, support, and a general competence score. Under such circumstances, the diagnostic value of subscale performance is greatly diminished.

Holistic scoring is usually simpler and faster than analytic; an important concern when teacher time is involved. Unless assessment's purpose is not to provide data to guide program improvement, a quick overview of achievement may be particularly suitable for program evaluation, for flagging students who need more help, and for assigning final evaluations.

Concurrent use of analytic and holistic strategies can optimize both diagnostic value and efficiency. One approach emerging from minimum competency testing is to score all essays holistically then rate analytically those essays that were scored below minimum competency. Another strategy, used in the Maine statewide assessment, is to score essays holistically, but to note analytic dimensions that are particularly strong or weak in an individual's work as a kind of generic "comment" on the performance.

Opinions differ considerably regarding the value of these different approaches, and research is ongoing. The important point is not so much the correct labeling of scales, but that a variety of approaches exist and can prove useful.

### What About Portfolio Assessment?

Portfolio assessment is often the first strategy that comes to mind when people think of alternative assessments. In some respects, portfolio assessment is a misnomer for "assessment of a body of work." In other instances, the portfolio assessment is really the assessment system. Portfolios are collections of student work that are reviewed against criteria in order to judge an individual student or a program. The portfolio or collection of work does not constitute the assessment; it is simply a receptacle for work (essays, videotapes, art, journal entries, and so on) that may or may not be evaluated. The "assessment" in portfolio exists only when (1) an assessment purpose is defined; (2) criteria or methods for determining what is put into the portfolio, by whom, and when, are explicated; and (3) criteria for assessing either the collection or individual pieces of work are identified. Deciding what should be included is really a task description, not a scoring guideline problem. What goes in, who chooses, when samples are taken—these are dimensions of the assessment task that define the setting and kinds of work that will be considered.

---

*You may be familiar with the term "Primary Trait Scoring." When Primary Trait criteria focus on only one trait, they are holistic; when expanded to two or more traits, they become analytic.

There are two issues related to selecting the dimensions of scoring criteria for portfolio assessment: (1) What are the criteria for selecting the samples that go into the portfolio, and (2) What are the criteria for judging the quality of the samples? Prior to considering criteria for judging portfolios, you will need to determine whether the portfolio should be rated as a whole or as individual samples. Second, you need to decide which dimensions reflect the intent or purpose of your assessment. When looking at a body of work, many issues arise, for example:

- Will progress or improvement be assessed?
- How or will progress be evaluated?
- How will different tasks, videos, art work, essays, journal entries, and the like be compared or weighted in the assessment?
- What is the role of student reflection in the assessment? Parental input?

Once these issues are settled, defining the dimensions of portfolio scoring criteria is the same as defining multidimensional criteria. Perhaps the best known example of portfolio assessment criteria is provided by the Vermont Mathematics portfolio, which is summarized in Figure 10. A body of mathematics work is evaluated on two major dimensions, problem-solving and communication skill. Within each dimension, several subdimensions further define each of the larger skills. Ratings are given for the subskills under the two dimensions, problem solving and communication. You can see how this example of portfolio assessment criteria resembles the multidimensional examples in Figures 1 and 7.

## Developing and Evaluating Scoring Criteria

### Beginning the Development Process

The process for developing your own criteria is straightforward:

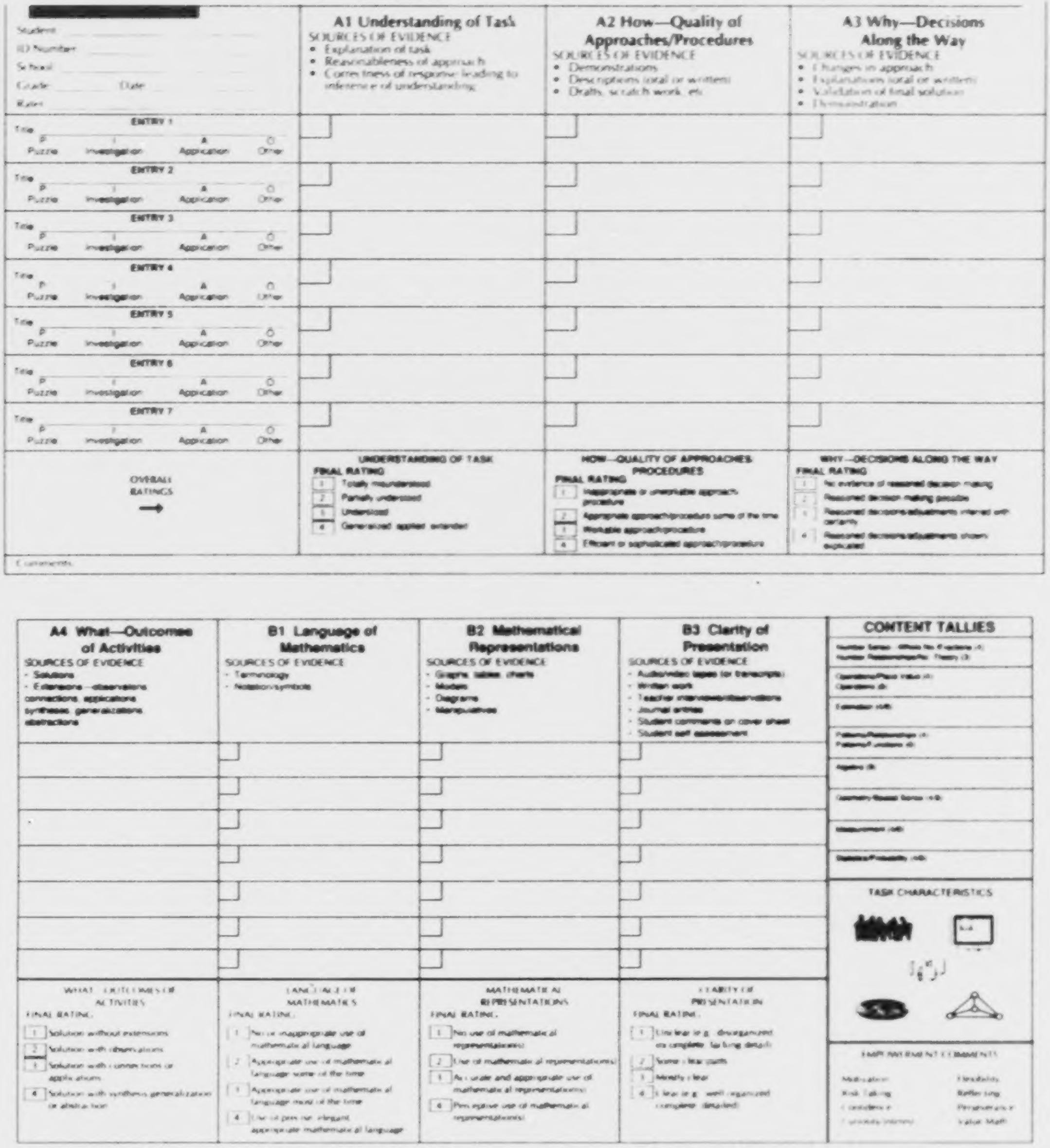
- Investigate how the assessed discipline defines quality performance.
- Gather sample rubrics for assessing writing, speech, the arts, and so on as models to adapt for your purposes.
- Gather samples of students' and experts' work that demonstrate the range of performance from ineffective to very effective.
- Discuss with others the characteristics of these models that distinguish the effective ones from the ineffective ones.
- Write descriptors for the important characteristics.
- Gather another sample of students' work.
- Try out criteria to see if they help you make accurate judgments about students.
- Revise your criteria.
- Try it again until the rubric score captures the "quality" of the work.

You probably noticed how recurrent this development process is. Initial ideas about important and scorable aspects of student performance become refined through use. Your criteria may focus on process—how a student approaches and solves a problem—as well as on the product or outcomes.

For example, we can refer to the development process for the criteria in Figure 5 (Baker, Aschbacher, Niemi, and Sato 1992). CRESST developed its rubric for rating depth of content understanding in history by collecting and examining the differences in essays written by history experts (university professors and graduate students in history) versus those written by novices (high school students). CRESST researchers looked for dimensions that seemed to differentiate the performance of these two groups. In a number of subject areas, the researchers observed differences between the students and the experts in the application of prior knowledge, the use of organizing concepts and principles, and misconceptions. These traits defined the first draft of scoring criteria. The criteria were then tried out on samples of student work and further clarified and refined to ensure that the scales were clearly defined, were appropriate for the range of student responses likely to be encountered, and enabled teachers or other raters to distinguish between essays that deserved adjacent points on a scale.

While undertaking the task of developing criteria, don't forget to take advantage of others' work. Quite often you can import or modify criteria from state and local assessment programs, curriculum experts, or colleagues who have grappled with similar assessment problems. Research literature on alternative assessment also provides examples of pilot alternative assessments similar to the one appearing in Figure 4, which can be adapted for classroom use. There is also a small but growing literature on the nature of expertise in various disciplines, such as how an historian reads and uses primary source documents.

## Figure 10
## Mathematics Rating Form



[redacted]

Student
ID Number
School
Grade Date
Rater

**A1 Understanding of Task**
SOURCES OF EVIDENCE
- Explanation of task
- Reasonableness of approach
- Correctness of response leading to inference of understanding

**A2 How—Quality of Approaches/Procedures**
SOURCES OF EVIDENCE
- Demonstrations
- Descriptions (oral or written)
- Drafts, scratch work, etc.

**A3 Why—Decisions Along the Way**
SOURCES OF EVIDENCE
- Changes in approach
- Explanations (oral or written)
- Validation of final solution
- Demonstration

ENTRY 1
Title
P Puzzle I Investigation A Application O Other

ENTRY 2
Title
P Puzzle I Investigation A Application O Other

ENTRY 3
Title
P Puzzle I Investigation A Application O Other

ENTRY 4
Title
P Puzzle I Investigation A Application O Other

ENTRY 5
Title
P Puzzle I Investigation A Application O Other

ENTRY 6
Title
P Puzzle I Investigation A Application O Other

ENTRY 7
Title
P Puzzle I Investigation A Application O Other

OVERALL RATINGS →

UNDERSTANDING OF TASK
FINAL RATING
1 Totally misunderstood
2 Partially understood
3 Understood
4 Generalized, applied, extended

HOW—QUALITY OF APPROACHES/PROCEDURES
FINAL RATING
1 Inappropriate or unworkable approach/procedure
2 Appropriate approach/procedure some of the time
3 Workable approach/procedure
4 Efficient or sophisticated approach/procedure

WHY—DECISIONS ALONG THE WAY
FINAL RATING
1 No evidence of reasoned decision making
2 Reasoned decision making possible
3 Reasoned decisions/adjustments inferred with certainty
4 Reasoned decisions/adjustments shown/explained

Comments

**A4 What—Outcomes of Activities**
SOURCES OF EVIDENCE
- Solutions
- Extensions—observations, connections, applications, syntheses, generalizations, abstractions

**B1 Language of Mathematics**
SOURCES OF EVIDENCE
- Terminology
- Notation/symbols

**B2 Mathematical Representations**
SOURCES OF EVIDENCE
- Graphs, tables, charts
- Models
- Diagrams
- Manipulatives

**B3 Clarity of Presentation**
SOURCES OF EVIDENCE
- Audio/video tapes (or transcripts)
- Written work
- Teacher interviews/observations
- Journal entries
- Student comments on cover sheet
- Student self assessment

WHAT—OUTCOMES OF ACTIVITIES
FINAL RATING
1 Solution without extensions
2 Solution with observations
3 Solution with connections or applications
4 Solution with synthesis, generalization or abstraction

LANGUAGE OF MATHEMATICS
FINAL RATING
1 No or inappropriate use of mathematical language
2 Appropriate use of mathematical language some of the time
3 Appropriate use of mathematical language most of the time
4 Use of precise, elegant, appropriate mathematical language

MATHEMATICAL REPRESENTATIONS
FINAL RATING
1 No use of mathematical representation(s)
2 Use of mathematical representation(s)
3 Accurate and appropriate use of mathematical representation(s)
4 Perceptive use of mathematical representation(s)

CLARITY OF PRESENTATION
FINAL RATING
1 Unclear (e.g., disorganized, incomplete, lacking detail)
2 Some clear parts
3 Mostly clear
4 Clear (e.g., well organized, complete, detailed)

**CONTENT TALLIES**
Number Sense - Whole No. to Fractions (4)
Number Relationships/No. Theory (8)
Operations/Place Value (4)
Operations (8)
Estimation (4-8)
Patterns/Relationships (4)
Patterns/Functions (8)
Algebra (8)
Geometry/Spatial Sense (4-8)
Measurement (4-8)
Statistics/Probability (4-8)

TASK CHARACTERISTICS

EMPOWERMENT COMMENTS
Motivation
Risk Taking
Confidence
Curiosity/Interest
Flexibility
Reflecting
Perseverance
Value Math

BEST COPY AVAILABLE

## Evaluating Criteria

Your criteria for judging students' work shape the decisions you eventually make about programs and students. Regardless of whether you are developing your own criteria or using those provided by others, it is important to review the quality of the scoring guidelines. We conclude this chapter with a proposed set of "criteria for criteria"— a checklist you can use to rate the quality of scoring criteria you borrow or develop. Our proposed criteria appear in Figure 11.

Now let's look at a set of dimensions for assessing the worth of your own criteria.

### Keyed to Important Outcomes

At a minimum, criteria for judging student performance need to address all the student outcomes you are trying to measure. For example, your criteria for judging student drama productions should encompass all the important drama and art that you want to be able to assess, and no others. If originality and logical presentation are part of the desired outcomes, you will want to include scales for judging these aspects of student work. If they are not an important outcome, omit them.

### Sensitive to Purpose

What educational decisions will you make on the basis of your assessment? The answer to this question should guide your decisions about whether to use a checklist or rating scales, how many scales, which traits, what types of scale, and so forth. Do you need a global, holistic view of student achievement or an analytical one that gives you information about several specific aspects of students achievement? Do you need the information in the form of a number for ease of reporting and aggregation at the expense of detail, or do you need the richness of qualitative description, or perhaps both?

### Meaningful, Clear, and Credible

The criteria by which you judge a performance need to be meaningful to students, parents, raters, teachers, administrators, policymakers, and the public. If the criteria are not credible, the results will probably be ignored or may be misused. Examples of student work that illustrate criterion traits can help make the criteria concrete for others. Involving others in the development of criteria increases their credibility.

Because one of the tenets of performance assessment is public and discussed criteria, your criteria need to make sense to students so that they will be able to apply them easily to their own work and become self-regulated learners. Although judgments of student performance tend to be subjective by their nature, they are more reliable and credible when they rely less on high inference and more on observable, concrete characteristics.

**Figure 11**
**How Do You Evaluate Scoring Criteria?**

- ☐ All important outcomes are addressed by criteria.
- ☐ Rating strategy matches decision purpose: holistic for global, evaluative view; analytic for diagnostic view.
- ☐ Rating scale provides usable, easily interpreted score.
- ☐ Criteria employ concrete references, clear language understandable to students, parents, other teachers.
- ☐ Criteria reflect current conceptions of "excellence" accepted in the field.
- ☐ Criteria have been reviewed for developmental, ethnic, gender bias.
- ☐ Criteria reflect teachable outcomes.
- ☐ Criteria are limited to feasible number of dimensions.
- ☐ Criteria are generalizable to other similar tasks or larger performance domain.

BEST COPY AVAILABLE

### Fair and Unbiased

Not only do assessment tasks need to be fair, but so do the criteria by which you define excellence. Unrecognized biases can seep into your definitions of traits, your specifications for what kind of performance earns which scale point, and your application of those criteria to individual pieces of student work. When you want your criteria to have diagnostic value, they must be sensitive to instruction and students' opportunities to learn the skills that are assessed. In contrast, you do not want them to reflect variables over which educators have no control, such as a child's culture, sex, or socioeconomic background.

### Feasible

Several reasons exist to limit the number and complexity of the performance dimensions to be judged. First, the time, effort, and money available for judging performance are always limited, sometimes severely so. Second, raters find it difficult to address too many different aspects of a work at once. In our experience at CRESST, raters were frustrated when asked to use more than six or seven scales for rating student essays. It became an onerous task and a less reliable process. Third, students will probably find it difficult to deal with too many aspects of their work at once. And finally, administrators and policymakers usually need information in as brief a form as possible. Separate scores for a large number of traits or for complex characteristics may make it more difficult to use the results effectively.

### Generalizable

Although we recognize that criteria for performance are strongly linked to discipline-based notions of excellence, rating can be more efficient when a single set of "generic" criteria can serve multiple topics, tasks, or disciplines. For example, we could develop a common set of criteria for assessing student understanding of science concepts through journals, hands-on experimentation, computer simulation, and oral presentation. We could also use a common set of criteria for judging student essays in social studies, science, and math? As disparate as these situations may seem, it is possible to envelop generic criteria for some purposes. If we could conceptualize excellence in consistent ways across assessment methods and disciplines, our criteria could have a more powerful impact on learning and instruction. Our example of the CRESST history-social studies rubric (Figure 5), which has also been applied to science and economics, shows one strategy for developing cross-discipline criteria. Like all good criteria, these proposed dimension are subject to revision and refinement.

### REFERENCES

Baker, E.L., P.R. Aschbacher, D. Niemi, and E. Sato. (1992). *CRESST Performance Assessment Models: Assessing Content Area Explanations*. Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Baxter, G., R.J. Shavelson, S. Goldman, and J. Pine. (Spring 1992). *Journal of Educational Measurement* 29, 1: 1-17.

Jones, E., and J.M. Roberts. (1990). *Profile of Developmental Outcomes for Kindergarten Literacy and Numeracy Skills*. Soledad, Calif.: San Vincente School District.

Vermont Department of Education. (1992). *Looking Beyond "the Answer": The Report of Vermont's Mathematics Portfolio Assessment Program, Pilot Year 1990-91*. Montpellier: Vermont Department of Education.

# VI-5

# What Is a Rubric? A Dialogue on Design and Use

GRANT WIGGINS
THE CENTER ON LEARNING, ASSESSMENT, AND SCHOOL STRUCTURE

A basic tool in a performance assessor's kit is the rubric. Rubrics specify the elements of performance that matter most and detail how to distinguish strong performances from weak ones. Questions abound concerning the design and use of rubrics, however: Should they be generic or task specific, holistic or analytic-trait? For whom are rubrics written—the novice performer or the veteran judge? Can a rubric yield reliable scoring without yielding valid inferences? How does one go about constructing a rubric? Let's explore these and other questions about rubric design and use.

**Just what is a rubric and why do we call it that? (Defining *rubric, criteria, indicator, and descriptor.)***

A **rubric** is a printed set of guidelines that distinguishes performances or products of different quality. For example, Diagram 1 shows a rubric for scoring a persuasive essay. A rubric has descriptors that define what to look for at each level of performance. In Diagram 1, the **descriptors** for the highest level include points a through d.

Rubrics also often have **indicators** providing specific examples or tell-tale signs of things to look for in work. In Diagram 1, an indicator for a level 6-5 performance would be "doesn't use enough phrasing like 'on the one hand . . . on the other hand.' "

The word rubric derives from the Latin word for red. Long ago, a rubric was the set of instructions for a law or liturgical service, typically written in red. Thus, a rubric instructs people on how to "lawfully" judge a performance. A good rubric allows valid and reliable—criterion-referenced—discrimination of performances.

**Help me with some of the other language often used in this work. What are criteria and how do they differ from standards? (Defining *criteria, standards, and benchmarks.)***

**Criteria** are the conditions a performance must meet to be successful. They define what it means to meet the task requirements. In Diagram 1, one of the criteria is coherence: "overall organization, and consecutiveness of thought." Think of the high jump and its criteria: a legal jump is up and over the bar without knocking it off. If the bar wobbles, the criteria are still met. Any high jump must pass that test to be successful.

---

Reprinted with permission from the author. Copyright © 1995 by Grant Wiggins. All rights reserved.

### DIAGRAM 1:
### WHAT IS A RUBRIC?
### HERITAGE HIGH SCHOOL (LITTLETON, COLORADO) SENIOR ESSAY
### (USING A READING PROMPT AND SCORING RUBRIC FROM THE COLORADO STATE PLACEMENT EXAM)

9-8 The upper range responses satisfy the following criteria:

a. **Summary.** The summary should identify the main idea (of the reading).

b. **Focus of agreement.** Agreement and/or disagreement may be complete or partial but writer must make clear what he/she is agreeing/disagreeing with. Specifically, 9-8 papers must address author's thesis, not substance abuse generally.

c. **Support for agreement/disagreement.** Support should provide an analysis of argument and/or relevant and concrete examples.

d. **Style of coherence.** These papers demonstrate clear style, overall organization, and consecutiveness of thought. They contain few repeated errors in usage, grammar, or mechanics.

(The four phrases in bold represent the dimensions being scored.)

(Two of the criteria are underlined.)

7 This grade is used for papers which fulfill basic requirements for the 9-8 grade but have less development, support or analysis.

6-5 Middle range papers omit or are deficient in one of these four criteria:

a. Summary. Summary absent or incomplete, listing only author's thesis.

b. Focus of agreement/disagreement. What the writer is agreeing/disagreeing with is not clear or unrelated to author's proposals. Example: writer doesn't use enough phrasing like "on the one hand . . . on the other hand . . ." (an indicator).

c. Support. Writer only counter-asserts; examples are highly generalized or not distinguishable from examples in the article. Analysis may be specious, irrelevant or thin.

d. Style and coherence. These papers are loosely organized or contain noticeable errors in usage, grammar, or mechanics.

4 This grade is used for papers slightly weaker than the 6-5 papers. Also, a student who writes his/her own parallel essay in a competent style should receive a 4.

3-2 These papers are deficient in two or more of the criteria. Typically, these papers weakly paraphrase the article OR they have serious organization/coherence problems. Papers with serious, repeated errors in usage, grammar or mechanics must be placed in this range. (This whole paragraph, like all the previous ones, is a descriptor for this point on the scale.)

A **standard** is different from criteria. In the high jump it tells us how high a successful jump must be to be considered excellent. Merely getting over the bar is not enough, especially if the bar is very low. How well did we meet the criteria? That's the "standards" question. Is three feet okay, or does the jump need to be six feet? In the high jump example, the standard is typically established in terms of specific heights that either need to be jumped or have been jumped by the best jumpers. It is based on patterns of actual performance and instances of exemplary performance.

Standards are set once the task and its appropriate criteria are established. More precisely, criteria are *inferred* from the goal and the nature of the performance while standards are *selected* to represent excellence in performance, using appropriate exemplary performances as benchmarks.

Consider a writing assessment. The criteria to be met typically involve things like clear prose, effective voice, and insightful analysis. The criteria are met or unmet on a continuum—from fully to incompletely. The standard states how well the criteria must be met in order for the performance to be considered good. *How* clear? *How* insightful?

Be careful, though; people use the word *standards* differently. It could mean the best there is, exemplary performance, the ideal; or it could mean minimal standards or "cut" scores. Thus, in Diagram 1, students must get a 4 to graduate. Often, when we ask whether a student's work is up to standard, we mean acceptable as opposed to exemplary. I prefer using the term to mean the best, as in Michael Jordan sets the world-class standard for shooting a basketball.

We make the same distinctions in writing at all levels: there are world-class professional standards set

BEST COPY AVAILABLE

by published authors; there are exit-level high school standards (and minimum standards); and there are grade-level standards. In each case, no matter how large or small the frame of reference, the standard is set by the best performances—the exemplars.

**Why isn't the top meeting of criteria the standard? I am confused. (Difference between standards and criteria.)**

A necessary criterion is different from a judgment about how well one must meet the criteria. *All* jumpers must jump over the bar; how high we want them to jump is the standards question. *All* papers, to be judged coherent, must meet the criterion of "overall organization, and consecutiveness of thought"; the level of organization depends on our standards. In the case of writing assessments, our standard for grade 3 is represented by "anchor papers," samples of the highest scoring work that best meet the criteria for grade 3.

We don't always know what the best looks like, however, particularly in classroom assessment, where the anchors are usually drawn from local student papers, and rubrics are derived from reflection on our experience. How do we know that our high-scoring local papers—our local standards—are truly excellent? We don't, unless we can relate them (through samples or correlations) to regional, state, and national standards. The criteria must always be met for the work to be considered acceptable. Writing should always be organized, clear, and insightful no matter who is doing it and in what context. The standards question is different—how good is good enough at those criteria in a particular context? Criteria usually stay the same; standards vary by context. For instance, an excellent high jump in high school probably does not equal an excellent high jump in Olympic competition. Again, remember that the ideal standard is different from the minimal standard or "cut" score, which establishes the line between meeting and not meeting the criterion.

**How is it that cut scores vary while the criteria stay the same? Why not just say that all papers that are "clear" are up to standard?**

Unlike the high jump, meeting intellectual criteria is a matter of degree and context. The value of the scores we give to any paper depends on our expectations and on the samples that anchor the top point on the scale. An exemplary 3rd grade "clear" paper looks unclear when measured against a professional writer's sample of clarity. If we use professional examples of clear writing as anchors, 3rd grade students will score very low and appear not to ever meet the criterion. If we use age- and experience-appropriate samples of excellent, clear writing the scores will be higher and some students will meet the criterion of clarity. Look at the rubric in Diagram 2 and tell me whether the rubric should be used to judge problem-solving work by calculus students or arithmetic work by 5th graders.

### DIAGRAM 2:
### RUBRIC FOR OPEN-ENDED MATH PROBLEMs

**DEMONSTRATED COMPETENCE**

**Exemplary Response...Rating = 6**

*Gives a complete response with a clear, coherent, unambiguous, and elegant explanation; includes a clear and simplified diagram; communicates effectively to the identified audience; shows understanding of the problem's mathematical ideas and processes; identifies all the important elements of the problem; may include examples and counter-examples; presents strong supporting arguments.*

**SATISFACTORY RESPONSE**

**Minor Flaws but Satisfactory...Rating = 4**

*Completes the problem satisfactorily, but the explanation may be muddled; argumentation may be incomplete; diagram may be inappropriate or unclear; understands the underlying mathematical ideas; uses ideas effectively.*

**INADEQUATE RESPONSE**

**Unable to Begin Effectively...Rating = 1**

*Words used do not reflect the problem; drawings misrepresent the problem situation; fails to indicate which information is appropriate.*

*Source: California CAP math report, 1989*

**That's a trick question! The rubric is clearly appropriate for use at any level where problem solving is taking place.**

That's the point. The criteria remain pretty much the same, but our expectations will appropriately vary based on the standards we hold students to. If we anchor the scoring process with very sophisticated calculus papers, then no arithmetic student will likely get a score above 1—not because they don't know calculus, but because their problem-solving skills will likely be relatively crude and limited. The criteria remain the same: solve the problem with clear, appropriate logic and substitution. On the other hand, if we anchor the scoring system with excellent 5th grade papers, then all but a few very incompetent or careless calculus students will get a 6 or a 5.

BEST COPY AVAILABLE

**Why would we ever want to anchor performance with samples of work beyond the capacity of our kids?**

To educate them about real excellence. Don't confuse standards with expectations. Very often young athletes and musicians have their work judged against standards (models or samples) of professionals. Look at diving or gymnastics; look at music fairs or contests. Even a 1st grade assessment of the making of an audiotaped book reading may usefully employ a professional sample as the standard. We don't expect 1st graders to produce professional products. Thus, please don't translate standard-referenced scores into letter grades. That wouldn't make sense, just as it doesn't make sense to say that novices in Spanish or German must all get poor grades because they aren't fluent (the standard).

**The rubric defines criteria and specifies dimensions of performance. Are criteria and dimensions of performance different ways of saying the same thing? (Holistic versus analytical-trait rubrics; advantages of analytical-trait.)**

**Dimensions of performance** concern where judges look; **criteria** concern what they will look for in judging whether or not success occurred for that dimension. One common dimension to score in assessing writing, for example, is "organization." The criteria for that dimension typically involve phrases like "there is a clear structure; the thesis is apparent and developed in a logical way; there is a beginning, a middle, and an end with a conclusion."

Sometimes an assessment that scores different dimensions separately is called an analytic-trait scoring process. For example, the dimensions we use to assess writing might be ideas, organization, voice, word choice, sentence fluency, and conventions. There would be a rubric for each dimension. Many rubrics, on the other hand, are holistic: they do not identify separate dimensions of performance. Instead, they assess performance as a whole.

**Why should I worry about so many different dimensions of performance? Why not just use a simple holistic rubric and be done with it?**

Holistic scoring might sacrifice validity and reliability for efficiency. Validity requires that we look for the right things—the apt and different dimensions of performance. Reliability is threatened if different judges unwittingly apply different criteria as they form an impressionistic whole judgment.

Consider a common problem in teacher grading. The student gets only one grade for a subject. We do not typically see or learn the criteria or dimensions of performance that are involved in the grading. Johnny and Suzy both get a *B* for the year, but for vastly different reasons. In one case it relates to inconsistent test scores, in another, consistent scores but poor homework. No one except the teacher knows the reasons for the grade—certainly not the reader of the transcript. The same teacher might give the students different grades at different times for varying reasons.

The more explicit the rubric, the more consistent and reliable the scoring. What happens if we use a vague, holistic rubric and each judge is then free to highlight different criteria? One math assessment judge is totally turned off by a student's computational errors; another sees them as minor, given the overall quality of the proof in response to a problem. Each judge is unknowingly applying different criteria. Unless the criteria specifically state a degree of computational accuracy and provide rubrics for how to weigh this factor in judging the overall quality of the problem, a judge can easily be distracted by obvious errors or achievements that may not be central to what is being assessed.

Another reason to use analytic-trait systems that score different dimensions of performance separately is that independent performance variables are usually examined in the same descriptor. Consider the oral assessment rubric in Diagram 3. What should we do if the student makes great eye contact but fails to make a clear case for the importance of the subject? It is easy to imagine that on the separate dimensions

### DIAGRAM 3: ORAL ASSESSMENT HOLISTIC RUBRIC

**5 - Excellent:** The student clearly describes the question studied and provides strong reasons for its importance. Specific information is given to support the conclusions that are drawn and described. The delivery is engaging and sentence structure is consistently correct. Eye contact is made and sustained throughout the presentation. There is strong evidence of preparation, organization, and enthusiasm for the topic. The visual aid is used to make the presentation more effective. Questions from the audience are clearly answered with specific and appropriate information.

***3 - Good:*** The student describes the question studied and conclusions are stated, but supporting information is not as strong as a 4 or 5. The delivery and sentence structure are generally correct. There is some indication of preparation and organization. The visual aid is mentioned. Questions from the audience are answered.

**1 - Poor:** The student makes a presentation without stating the question or its importance. The topic is unclear and no adequate conclusions are stated. The delivery is difficult to follow. There is no indication of preparation or organization. Questions from the audience receive only the most basic, or no, response.

of "contact with audience" and "argued for importance of topic," a student might be good at one and poor at the other. The rubric has us believe that these achievements always go together. Logic and experience suggest otherwise.

**What do analytic-trait rubrics look like?**

Diagram 4 shows one for assessing seminar performance in a high school class. (To save space, only definitions for scores 4 and 1 are shown.)

### Diagram 4:
### ANALYTIC-TRAIT RUBRIC FOR ASSESSING PERFORMANCE IN A SOCRATIC SEMINAR

**EXCELLENT - 4**

**A. Conduct.** Demonstrates clear respect for the learning process and patience with different opinions and complexity. Shows initiative by asking others for clarification, bringing others into the conversation, and moving the conversation forward. Speaks to all participants (not just the teacher or a single student). Avoids talking too much, too long, too softly. Avoids nit-picking and inappropriate language.

**B. Speaking/Reasoning:** Understands questions before answering. Cites evidence from the text (or elsewhere when appropriate). Expresses complete thoughts in complete sentences. Comments are logical and insightful and move the conversation forward. Makes connections among ideas from previous speakers or resolves seemingly contradictory ideas. Considers all sources, not just his or her own. Avoids bad logic.

**C. Listening:** Pays attention to details, writes down questions and thoughts. Responses take into account the comments of other participants and demonstrates that the student has kept up. Points out bad logic. Overcomes distractions.

**D. Reading:** Student is thoroughly familiar with the text and has come prepared with notes and questions in the margins. Key words, phrases, and ideas are underlined and possible contradictions have been identified. Pronounces words perfectly.

**UNSATISFACTORY - 1**

**A. Conduct:** Displays little respect for the learning process. If participating at all, the student may be argumentative. Minor distractions are taken advantage of. May use inappropriate language and speak about individuals rather than ideas. Arrives unprepared without notes, a pencil, perhaps even the text.

**B. Speaking/Reasoning:** Extremely reluctant to participate even when called upon. Comments are so illogical as to be meaningless. May mumble or express incomplete ideas. Little or no account is made of previous comments or important ideas in the text.

**C. Listening:** Appears uninvolved in the seminar. Comments may display complete misinterpretation of questions or comments of other participants.

**D. Reading:** Student is clearly unprepared for seminar. Important words, phrases or ideas in the text are unfamiliar. There are no notes or questions marked on the text and no attempt has been made to get help with difficult reading.

Source: Courtesy of North Shore High School, Glen Cove, New York

**Breaking performance into independent dimensions is so useful. Why do people even bother with holistic scoring?**

Holistic scoring is quick, easy, and often reliable enough when we are assessing a generic skill like writing. It's a trade-off, a dilemma.

**What did you mean when you said that rubrics could affect validity? Why isn't that a function of the task as designated? (Considerations in designing rubrics: include the most important and revealing dimensions of performance.)**

Validity concerns what you can conclude from scores. Tests or tasks are not valid or invalid; inferences based on results are valid or invalid. Suppose a rubric for judging story writing places exclusive emphasis on spelling and grammatical accuracy. The scores will likely be highly reliable—because it is easy to count the errors—but yield invalid inferences about students' ability to write stories. It is unlikely that spelling accuracy correlates with the ability to write an engaging, vivid, and coherent story. Many fine spellers can't construct engaging narratives, and no doubt many wonderful storytellers did poorly in school grammar and spelling tests.

You should consider, therefore, the appropriateness and validity of not only the performance task, but also the rubric. Given the task and the capacities being assessed, are we focusing on the most apt criteria? Have we identified the most important and revealing dimensions of performance, given the criteria most suitable for such an outcome? Does the rubric provide an authentic and effective way of discriminating between performances? Are the descriptors for each level of performance sufficiently grounded in actual samples of performance of different quality? These and other questions lie at the heart of rubric construction.

**How do you answer such design questions?**

By focusing on the *purpose* or desirable *impact* of performance, not just the most obvious features of performers or performances. Too many rubrics focus on obvious successes or failures that may be incidental to whether the overall result or purpose was achieved. Judges of math problem solving, for example, could focus too much on obvious computational

BEST COPY AVAILABLE

errors; judges of writing could focus too much on syntactical or mechanical errors. We don't want to score only the obvious strengths and weaknesses—the ones that are easy to see and count. We want to use *criteria that relate directly to the purpose and nature of the task*.

**Are there any simple rules for better distinguishing between valid and invalid criteria?**

One simple test is negative: Can you imagine someone meeting all the proposed criteria but not being able to perform well at the task, given its true purpose or nature? If so, then you have the wrong criteria. For example, some writing rubrics assess organization, mechanics, accuracy, and appropriateness to topic in judging analytic essays. These criteria are necessary, but perhaps not sufficient. They don't get to the heart of the purpose of writing—achieving some effect or impact on the reader. A writer can meet these form-related or process-related criteria but still produce bland, uninteresting writing. Thus, they cannot be the best set of criteria or dimensions.

**But surely the formal and mechanical aspects of performance matter.**

Of course they do. But they don't necessarily get at the point of writing, merely the means of achieving the purpose. What is the writer's intent? What is the purpose of any writing? It should "work" or yield a certain effect on the reader. *Huck Finn* works even though the written speech of the characters is ungrammatical. The writing aims at a result. Writers aim to accomplish a response—that's what we must assess for. The real criteria will be found by analyzing the answers to questions about purpose. If we assess analytical writing, for instance, we should look for things like the insightfulness, novelty, clarity, and compelling nature of the analysis.

Or take speaking: Many good speakers make eye contact and vary their pitch, but those are not the criteria of good speaking; they are merely useful behaviors for achieving the intent. Criteria must relate to the purpose of the speaking, the desired *effect* of my speech. Was I understood? Was I engaging? Was I persuasive or moving? Whatever my intentions, did I realize them? Too many rubrics focus on useful behavior for the task (eye contact) and not enough on the purpose of the task (to interest and inform the audience) or the key demands of the situation and whether they were realized. The useful behavior may often occur in good performance, but it isn't the essence of good performance. We can easily imagine or find examples of good speaking without eye contact, and we can find examples of dreary speaking with eye contact. (Certain behaviors are called "indicators" because they tend to occur in good performance, but don't guarantee it.)

**You're saying we should not assess useful process or form traits of performance?**

No, I didn't mean to suggest it is always a mistake. Specific behaviors help us evaluate the quality of a performance. Rather, we should look at how processes and results work as a whole, develop *separate* rubrics, and weigh each dimension according to its importance in scoring the accomplishment. For example, in archery one might wish to score stance, technique with the bow, and accuracy. Stance matters. On the other hand, the ultimate value of the performance relates to its accuracy. In practice that means we can justifiably score for a process, but we shouldn't overvalue process so much that results don't matter much. Nor should we arbitrarily weigh processes and results equally, as too many rubrics do.

Weighing can vary over time. Accuracy may be worth only 25 percent when scoring a novice, but 100 percent when scoring an expert. In my experience, teacher-designed rubrics tend to overvalue processes and useful behaviors while undervaluing the quality of the result. That tells the student that obeying instructions is more important than succeeding.

**Given how complex this is, why not just say that the difference between the levels of performance is that if a 6 is thorough or accurate, then a 5 is less thorough, less clear, or less accurate. Most rubrics seem to do that; they rely on a lot of comparative and evaluative language. (Considerations in designing rubrics: go beyond scales using comparative language.)**

Alas, you're right. This is inadequate, in my view. To just use comparative language is lazy or novice practice that stems from a failure to keep seeking through analysis descriptors of the unique features of strong, middle, and weak performance. Maybe the set of samples simply has "less" clarity, accuracy, or whatever, but the best descriptors note *qualitative* features unique to each pile, not *quantitative* arbitrary distinctions.

The student is left with weak feedback when rubrics rely heavily on descriptors like "less than a 5" or "fairly complete performance"—this isn't much different from getting a paper back with a letter grade. Ideally, a rubric focuses on discernible and useful differences in performance; that way the assessment is educative, not just measurement.

Compare the following excerpts from the ACTFL guidelines (Diagram 5) and a social studies rubric (Diagram 6) to see the point. The ACTFL rubric is rich in descriptive language and provides insight into each level and its uniqueness. The social studies rubric never gets beyond comparative language for the dimensions being assessed. In the social studies rubric, note that the only difference between each score point is a change in one adjective.

### DIAGRAM 5
### FROM THE ACTFL SPANISH PROFICIENCY GUIDELINES

**Novice-High:** Able to satisfy immediate needs using learned utterances . . . can ask questions or make statements with reasonable accuracy only where this involves short memorized utterances or formulae. Most utterances are telegraphic, and errors often occur when word endings and verbs are omitted or confused. . . . Speech is characterized by enumeration, rather than by sentences. There is some concept of the present tense forms of regular verbs particularly -ar verbs, and some common irregular verbs. . . . There is some use of articles, indicating a concept of gender, although mistakes are constant and numerous. . . .

**Intermediate-High:** Able to satisfy most survival needs and limited social demands. Developing flexibility in language production although fluency is still uneven. Can initiate and sustain a general conversation on factual topics beyond basic survival needs. Can give autobiographical information. . . . Can sporadically, although not consistently, give simple directions and narration of present, past and future events, although limited vocabulary range and insufficient control of grammar lead to much hesitation and inaccuracy. . . . Has basic knowledge of the differences between ser and estar, although errors are frequent. . . . Can control the present tense of most regular and irregular verbs. . . . Comprehensible to native speakers used to dealing with foreigners, but still has to repeat utterances frequently to be understood by general public.

Yes, in the beginning of our work, we may need to lean on comparative and evaluative language. For a first rubric, being clear about the highest level of performance and then using comparative language to describe the progressive weakness of other levels may be okay. But we certainly need to refine the rubric as soon as we have lots of performances to analyze. It is never valid, however, to confuse comparative *qualitative* differences with the often arbitrary *quantitative* differences in discriminating between performances.

**What do you mean? (Considerations in designing rubrics: avoid quantitative scales.)**

A less clear paper is obviously not as desirable as a clear paper (even though that doesn't tell us much about what clarity or its absence looks like), but it is almost never valid to say that a good paper has *more* facts, footnotes, or arguments than a worse paper. A paper is never worse because it has fewer footnotes; it is worse because the sources cited are somehow less appropriate or illuminating. A paper is not good because it has "more" words, but because it has something to say. There is a temptation to construct descriptors based on quantities instead of qualities. We love to score what is easy to score, not necessarily what is apt.

A rubric should always detail better and worse in tangible, qualitative terms. What specifically makes this argument or proof better than another one? If you use comparative language to differentiate quality, make sure that what is being compared is relative quality, not relative, arbitrary quantity.

### DIAGRAM 6
### FROM A CANADIAN SOCIAL STUDIES ESSAY EXAM

(Rubric concerns quality of examples or cases used by author; other rubrics exist for quality of argument and historical accuracy.)

| SCORE | SCORING DESCRIPTORS |
|---|---|
| (5)<br>Exceptional | The examples or case studies selected are relevant, accurate and comprehensively developed, revealing a *mature and insightful* understanding of social studies content. |
| (4)<br>Proficient | The examples or case studies selected are relevant, accurate and clearly developed, revealing a *solid* understanding of social studies content. |
| (3)<br>Satisfactory | The examples or case studies selected are relevant and adequately developed but may contain some factual errors. The development of the case studies/examples reveals an *adequate* understanding of social studies content. |
| (2)<br>Limited | The examples or cases selected, while relevant, are vaguely or incompletely developed, and/or they contain inaccuracies. A *restricted* understanding of social studies is revealed. |
| (1)<br>Poor | The examples are relevant, but a minimal attempt has been made to develop them, and/or the examples contain major errors revealing a lack of understanding of content. |

**This is hard work. How does one design a good rubric? (How are rubrics designed? Steps in development versus what good rubrics look like.)**

Your question is ambiguous. Do you mean "by what process" or "against what criteria" are rubrics designed?

**Why do you ask? What's the difference?**

There is a great deal of difference. In one case, you want to know the best chronology of events in the design process. In the latter, you want to know the criteria against which your product of design—by whatever process it was designed—will be evaluated and modified.

Think of cooking. To follow a complicated recipe in a lock-step way without assessing it against a hoped-for result will typically yield an unsatisfactory product. Similarly, good cooks can vary the sequence, even the ingredients, and have the meal come out splendidly. They know and wisely apply the criteria by which the final dish will be judged. If not, they might burn the casserole, even though the oven *was* at the 425 degrees the recipe called for, and they *knew* the recipe called for 10 more minutes of cooking at that temperature. We determine whether such a challenge is met or not met through ongoing assessment against the criteria—taste, smell, or observation of the dish, keeping in mind the desired results. "I did what the recipe said" is as unacceptable an excuse in cooking as it is in the classroom. Too often, the student has no idea what the result should be; too often, assessments seek to learn if the recipe has been followed (e.g., write a five-page essay). Following the form seems to matter more than the quality of the result.

The distinction helps explain, by the way, why many outcome-based reform projects end up with sterile outcomes and assessments. People confuse following the logic of going from outcomes to criteria to tasks to standards with a recipe for how to come up with standards. The real test is whether we have good standards, not whether we followed all the steps. The logic of OBE was never meant to be a one-way chronology of how the work should be undertaken.

**What does this have to do with rubrics?**

The same problem occurs in rubric design as in thoughtless cooking or school reform: People either want a recipe to follow blindly, or they think that effective change depends on following a sequence instead of intelligently applying a method against feedback. Rubric construction is not like blindly following a checklist or recipe; it is more like turning a blueprint into a building. The blueprint provides the criteria against which all ongoing work is judged—and adjusted regardless of whether contractors disagree about a calendar for building.

**I see the difference, but what I wanted to know were some tips and guidelines about how to actually develop a good rubric. (How are rubrics designed? Analyze actual samples of performance.)**

Let me answer first in the negative. You do not design through mere imagination and discussion. You must derive rubrics by analyzing existing samples of performances of different quality, starting with examples of excellence. The rubric must reflect the most tangible and appropriate differences of quality between actual performances. We sort work into piles of decreasing quality. It's only when we stop to look at how the piles differ, and to carefully explain the differences, that we begin to get a clear sense of what the descriptors of the rubric are for each pile.

The first year, before you have adequate performances, you have to do the best you can, basing your rubric on your overall experience of differences in this type of performance, any piloting you did, or samples of performance on past *similar* challenges.

**So I should refine rubrics on the basis of future performance testing and results, regardless of how I handle the first year?**

Absolutely. The rubric is only as good as the breadth and depth of your sample of performances and your power to generalize the results of sorting the work into piles. Each year brings a more diverse and revealing sample of possible performances, hence deeper insight into the most salient differences in performance levels and the surprising diversity of work possible within each level.

**Then the best rubrics depend on a clear and uncontroversial definition of exemplary performance, as gained from exemplary samples, and one works down the scale from there.**

Precisely! That means we must have some faith that our samples of so-called exemplary performances really are exemplary, and that our criteria derive from an analysis of excellence, not just abstract ideas about excellence. This is a problem that strikes at the heart of the differences between norms and standards. Maybe we have only samples that are the best products our kids are capable of doing—a best that may not really be the best that is possible. We might then be building rubrics out of an analysis of mediocrity—describing traits of merely acceptable performance, but claiming them as the highest level of performance. Imagine, for example, if we derived the criteria of exemplary musicianship from the best performances of relative novices, and you'll see the problems we face if we begin with a poor range of work.

**Are criteria subjective? Matters of taste or value? Dependent on the designer's arguable sense of what constitutes good performance? (Commonly asked question: Are criteria subjective?)**

From one vantage point, criteria are not matters of value or taste; they are the result of analyzing the purpose and intended impact. If my purpose is to solve problems or write stories, then, by logical analysis, some criteria are apt and others aren't. This is as objective as saying that having two men on the same base is wrong in baseball. It's not a subjective, aesthetic, or moral judgment but a logical one; the rules and purpose of the game don't permit it. That doesn't mean that we won't disagree about a complex analysis of implications. But criteria should be logically derived from an analysis, purposes, and what is effective in meeting them.

Put differently, our choice of purposes may be a value question, but the identification of criteria, given a clear purpose and agreed-upon standards, is a logical question: Which criteria fit and make the most sense? We may disagree about the value of story writing, but we should be able to agree about the criteria for distinguishing good from not-so-good story writing once the exemplary samples are in place. We do that by asking, "What is the essence of the good stories we hold up as exemplary? When stories work,' why do they work?"

**Whoa! Isn't that a value question? Maybe I think the stories of Hemingway work and you don't.**

Then we disagree about the standards—the samples of exemplary writing. Once we agree about samples, the criteria are derived from an analysis of what we chose. You say Shakespearean sonnets are exemplary poems; I say the songs of Bob Dylan are. We disagree. Now what? We must work toward a common understanding of the criteria by which we define poems and their impact. Maybe Dylan and Shakespeare have some traits in common that are important, but not obvious, on superficial inquiry. That's why good judges are good; they typically see what novice observers can't. It's hard, messy work—with no recipes. Typically, disagreements occur over which criteria should matter most from among the many.

**I'm beginning to wonder if I like the messiness. Speaking of messiness, are you suggesting that all the criteria we value can be formalized into rubrics? What about such messy, hard-to-pin-down outcomes like creativity and open-mindedness? (Can all criteria be formalized into rubrics?)**

This is an important question and one that rarely gets the airing it should. My answer may not satisfy you, but it is consistent with what we said before. If we have samples of exemplary performance for the outcome in question, then we can derive a valid rubric. If we agree that this pile of work over here is more creative than all the other piles of work we have, then through hard analytical work we should be able to describe what it is that makes the work creative. I might go so far as to say that if we value it and recognize it when we see it, then we can assess the work with adequate precision. Think of the surprising accuracy in formal wine tasting or the artistic elements in the Olympics.

**Should a rubric identify only the strengths of each level? (Considerations in designing rubrics.)**

No. Often the errors are hallmarks of all lower levels of performance. Look back at the ACTFL foreign language proficiency levels (Diagram 5). They identify the most characteristic features of each level, including the most telling errors. Your question reminds me of another complication. Sometimes perfectly valid rubrics do not use descriptors that strike us as the essence of a level of performance.

**How can that be? (Considerations in designing rubrics: include essential points of differences in performances.)**

Judging requires us to create levels of valid, reliable performance discrimination. Judges learn that certain features of performance make accurate discrimination easier—even though the distinguishing features may strike us as less important. For example, it may be that the most important difference between the best persuasive essays and the next best batch has less to do with some overall quality of the argument than the minor errors that distract from the argument (see Diagram 5 for examples).

**But if the judges don't routinely look at essential points of difference, won't it be tougher for students and teachers to learn from performance and properly prepare for it?**

Yes, I think so. The student might well focus too much on editing or minor mistakes rather than on judging the piece and its effectiveness as a whole. The judge's tricks don't necessarily help the test-taker or teacher understand or practice for the ultimate challenge, the criterion performance. Here, as elsewhere, we must ask ourselves why, if the aim is merely accurate scores, we are assessing.

**Should a rubric be task-specific or more general? (Considerations in designing rubrics: use generic criteria.)**

Good question. Somewhere in between really. We face a dilemma. Reliability is served by having a rubric unique to each task. But we have a feasibility and maybe reliability problem: it takes too much time and energy to design a rubric for every task. Furthermore, a specially designed rubric for each task may not be necessary. For example, a rubric for evaluating a student's ability to "argue persuasively" and "master the core understandings of a text" may be suitable for specific topics or oral exams—particularly if the important outcomes (such as "argue persuasively") can be viewed as the main things to assess.

**Can't assessment artificially narrow the realm of possible performance? Won't the use of rubrics stifle creativity? (Commonly asked questions: will rubrics stifle creativity?)**

Sometimes rubrics can constrain performer imagination. Because some designers seem intent on creating rubrics by bland consensus, they may avoid controversial or difficult-to-assess characteristics of performance. Look at the rubric from NAEP in Diagram 7.

Note that one could write the most boring, uninteresting story in the world and still get a 6 if it fit the formula asked for by the rubric. Compare that approach to creativity with the one implied in an ad for the J. Walter Thompson advertising agency appearing on the back of the *New York Times* last year. Among the eight performance tests for would-be copywriters was this: "You are a writer for Walletsize Books. Describe the history of the United States in 100 words or less." The advertisement went on to say, "So if you want to get our attention and a job, you're going to have to show us fresh, fearless, and more or less brilliant stuff."

**Cute, but I'm not convinced. Some people say rubrics inhibit creativity. What do you say to those claiming that rubrics may lead to better overall performance, but they ultimately result in formulaic and restricted performance? After all, the Impressionists were deemed to be poor painters by the Academy in France.**

First of all, let me say that stifling creativity has less to do with rubrics per se than with using *inappropriate* rubrics and a single anchor performance. With single anchor performances, we encourage students to mimic a particular approach. As we noted earlier, the more we assess approach or format instead of the criteria of the accomplishment, the more we encourage mimicry. With samples of *diverse* excellence, we encourage the performer to consider novel, maybe even unorthodox, means of achieving the end.

We never want to build rubrics that call for orthodox or formal behavior; we should never mandate a process, format, or approach. As I said before, that mistake tends to happen only when we try to exclusively score form or methods instead of impact or accomplishments.

### DIAGRAM 7
### NARRATIVE SCORING GUIDE

1. **Event Description.** Paper is a list of sentences minimally related or a list of sentences that all describe a single event.

2. **Undeveloped Story.** Paper is a listing of related events. More than one event is described, but with few details about setting, characters, or the events. (Usually there is no more than one sentence telling about each event.)

3. **Basic Story.** Paper describes a series of events, giving details (in at least two or three sentences) about some aspect of the story (the events, the characters' goals, or problems to be solved). But the story lacks cohesion because of problems with syntax, sequencing, events missing, or an undeveloped ending.

4. **Extended Story.** Paper describes a sequence of episodes, including details about most story elements (i.e., setting, episodes, characters' goals, problems to be solved). But the stories are confusing or incomplete (i.e., at the end the characters' goals are ignored or problems are inadequately resolved; the beginning does not match the rest of the story; the internal logic or plausibility of characters' actions is not maintained).

5. **Developed Story.** Paper describes a sequence of episodes in which almost all story elements are clearly developed (i.e., setting, episodes, characters' goals, or problems to be solved) with a simple resolution of these goals or problems at the end. May have one or two problems or include too much detail.

6. **Elaborated Story.** Paper describes a sequence of episodes in which almost all story elements are well developed (i.e., setting, episodes, characters' goals, or problems to be solved). The resolution of the goals or problems at the end are elaborated. The events are presented and elaborated in a cohesive way.

*Source:* National Assessment of Education Progress (NAEP), Portfolio Assessment, 1992.

**Wait a minute! Shouldn't we require things like the scientific method?**

Many people are confused about this. As noted above, the method is not a fixed set of steps to be taken in a certain order, but criteria to be met. It doesn't matter how you begin as long as you are methodical in isolating variables. Yet when many science teachers design a problem-solving rubric, they err by requiring students to use *the* five-step procedure they teach in their classes. Dewey warned about this years ago in *How We Think*. He repeatedly urged readers not to think of the five stages of reflective thinking as a series of invariant steps but as a set of criteria. That didn't prevent countless researchers from ossifying his criteria into rules of procedure.

Methods are means to ends. The rubric should refer to the judging of both, with emphases on the end. If the goal is to build a working roller coaster or debate a position, we don't typically need and shouldn't specify a required method. We need to specify the characteristics of the roller coaster or debate. Creativity is heightened, not impeded. Unorthodox approaches will yield high scores if they work—that is, if they thereby yield the results sought.

The basic method for avoiding this problem is as follows: First, anchor your scoring with diverse, even divergent, models of excellence. You might choose Shakespeare, Auden, and e. e. cummings samples for poetry, or Jesse Jackson and Ronald Reagan videotapes for public speaking. In doing so, you make clear that there is no one method, style, or approach needed to meet the objective, and thus focus students on purpose and impact. You make it more likely, not less, that students will invent novel approaches to a task. As I have said throughout, rubrics depend on clarity about standards and purposes.

**But that only addresses half my concern. The French Academy thought that Manet and the others were not meeting the criteria of art. They could be said to have defined art and its purpose in such a way that Impressionism was ruled out as an appropriate result.**

No, I think they disagreed about purpose and method. When you put it that way, I agree there is a danger. Maybe we should say that what changed were conventions and tastes about which of the *possible* criteria should be most heavily weighted. Jazz, rock, and rap are now seen as music, not because the purpose of the creators or judges changed, but because we have downplayed melody and formal structure in favor of other criteria—rhythm and the importance of structured spontaneity. We often disagree about which criteria should carry the most weight, even if we agree on all the criteria.

**But that way of putting it downplays what many fear most: That an orthodoxy of taste or criteria will hamper performance and progress. Couldn't new criteria enter into our thinking? Don't we have to leave open that possibility?**

Of course criteria will evolve—as our sense of purpose evolves. Look at something as simple as basketball. It has evolved into a contact sport, and both the criteria and our use of them in judging personal fouls have changed accordingly (hand checking used to always be illegal, for example). If we let this problem stop us from defining what we currently value in work, then we miss an opportunity to open discussion with colleagues and students that will allow us to reach common current understandings and explore how criteria evolve with insight and experience. Criteria are there whether we acknowledge them or not. Why not make their exploration a public process? Only a feeling for precise and timeless scoring procedures prevent it.

**What are some other unresolved issues in this move toward greater judgment-based assessment using rubrics? (Other issues: who does the rubric serve?)**

One issue concerns the basic purpose and audience: Who is the rubric for and what is it meant to accomplish? If the rubric is for the judge, then the descriptors will likely focus exclusively on very technical elements, strengths and weaknesses of the performance or product. But for the richest and most useful diagnostic feedback, the performer needs guidance about the process strengths and weaknesses and how the weaknesses might be corrected. We don't want the student to think that eye contact and note taking are *essential* to good listening, but if we think that their performance suffers because of inattentiveness, then we might want to recommend more eye contact and note taking. How much of the rubric should contain language about process isn't something that's set in stone. It certainly is the job of the teacher or coach, however, to supply diagnostic help.

**Most rubrics seem to focus on mastery of skills or processes such as writing or speaking. Are rubrics appropriate for content mastery? (Other issues: assessing content knowledge.)**

Yes, but you wouldn't know it from much of the work going on. Most educators choose to focus on the assessment of generic skills instead of specific subject matter understandings, but there's no technical reason to do this. It is certainly *easier* to assess general intellectual skills, and the assessment results usually allow for greater generalizability. Speaking is speaking, after all, even if the content of the speech varies. But understanding F=ma has nothing to do with understanding Boyle's Law or the concept of dramatic irony. Few teachers or assessors have sufficiently explored the question of what counts as evidence of *understanding* F=ma (as opposed to thoughtless control over the formula and plug-in use of it in typical textbook problems) and how we should distinguish between crude and sophisticated understandings. Typically, all that is assessed is knowledge and application of a formula, in Bloom's sense of those terms (note that "application" for Bloom is not a higher-order thinking skill).

**Can we really distinguish between levels of understanding in the same way we can levels of skill?**

Of course we can. You're forgetting the basic rule: If we have samples of work that represent a range of performance, we can develop rubrics (Piaget did it for decades). Maybe one reason you don't see this as easily is because we tend to think in terms of right or wrong answers only when dealing with subject-matter content. We tend to think of assessment of content control in terms of accuracy of facts and application. This is a mistake, and an easier one to see if we use words like "depth of insight" or "sophisticated" versus "simplistic" in describing different understandings. If we start to think that way, we can then talk about gradual differences in understanding. We could chart a student's insight on a scale that has "deep and sophisticated insight" at one end and "crude and superficial understanding" at the other. Take a look at a 5th grade paper on the causes of the Civil War and compare it with a fine 11th grade paper to see the point.

**Could you sum up for me what we've talked about? (Summary)**

Sure. Here are a few general principles and guidelines, based on what we said.

1. A rubric:

- Contains a *scale* of possible points for scoring work on a continuum. High numbers usually are assigned to the best performances. Scales typically use numbers from 4 to 8 as the top score and the numbers 1 or 0 for the lowest score.
- Provides *descriptors* for each level of performance assessed.
- Is either *holistic* or *analytic*. If holistic, the rubric has only one general descriptor for the performance as a whole. If analytic, there are multiple rubrics corresponding to each dimension of performance or trait being scored.
- When analytic-trait, the rubric provides guidelines for assessing the most *feasible and apt dimensions* of performance being examined. It breaks performance into major independent sub-scores (e.g., "syntax" and "voice" in writing; "precision of calculation" and "understanding of scientific method" in science).

2. The rubric must enable judges and performers to *effectively discriminate* between performances of different quality.

a. The discrimination must be valid: The dimensions being assessed and the characteristic differences of each level of performance must be salient, not arbitrary.

b. The discrimination must be reliable: The scores yielded by either the same judge at different times or different judges at the same time must be consistent within reasonable limits.

- Rubrics, and tests in general, tend to sacrifice validity for reliability for easily obtained, stable, and defensible scores. Beware!

3. The descriptors should use language that *maximally describes* each level of performance and its most salient and defining characteristics.

a. For the rubric to contain rich descriptors of levels of performance, the descriptors must be generalizations derived from *actual samples* of student work.

b. Rubrics are likely to sacrifice validity for reliability when relying heavily on *comparative or evaluative language* for the discrimination instead of performance- or behavior-focused descriptors.

- Typically, the student is improperly rewarded or penalized for the *quantity* of work instead of the *quality* of work.

c. Rubrics are meant to yield criterion-referenced results, not norm-referenced results. Thus, judges *cannot predict* the spread of scores. We would not even expect a "normal" curve.

4. Because rubrics are meant to be criterion-referenced, the highest point on the scale should describe exemplary performance as derived from samples of genuine excellence.

a. *Standards are not the same as expectations.* Scoring should alert the students to their real levels of performance. It may well happen that no one gets the highest score; it may happen that many students get low scores.

5. The most important points on the rubric scale are the top point and the "cut" point—scores that distinguish passing from failing.

a. Until you anchor the top point on the scale with genuinely excellent samples, the test is likely to be norm-referenced—based on who does the best work locally.

b. It is essential to validate the cut score carefully to ensure that all constituencies find the assessment process fair and valid.

- We cannot declare what constitutes a passing score before looking at work samples. It does not follow that a 2 on a five-point scale is passing merely because we call a *D* a passing grade on our five-point letter grade system—unless the faculty judges level 2 work samples to be just passing and level 1 work unacceptable.

6. The number of points on the scale should be big enough to be effective, but small enough to be reliable and manageable. Research and common sense suggest that the optimum number is somewhere between 4 and 10 points on a scale for use in local assessment. An even number of points is better than an odd number because it forces more care in the judging.

7. The variation between descriptors of each point on the scale should represent a smooth continuum.

a. The performance variability between each point on the scale and its descriptors should be smooth across all points (i.e., descriptors for a 5, 4, and 3 should not be excellent, good, and poor. The jump between good and poor is larger than the jump from excellent to good).

8. The more task-specific the rubric, the more valid the result; the more clear and simple the rubric, the greater the reliability.

## RECOMMENDED READINGS

American Council on the Teaching of Foreign Languages. (1982). *ACTFL Provisional Proficiency Guidelines*. Hastings-on-Hudson, N.Y.: ACTFL Materials Center.

Archibald, E., and F. Newmann. (1988). *Beyond Standardized Testing: Authentic Academic Achievement in the Secondary School*. Reston, Va.: NASSP Publications.

Baker, E., et al. (1991). *Cognitively Sensitive Assessments of Student Writing in the Content Areas*. Los Angeles, Calif.: The National Center for Research on Evaluation, Standards and Student Testing (CRESST).

Brandt, R., ed.(1992). *Performance Assessment: Readings from Educational Leadership*. Alexandria, Va.: Association for Supervision and Curriculum Development.

California State Department of Education. (1989). *A Question of Thinking: A First Look at Students' Performance on Open-Ended Questions in Mathematics*. Sacramento, Calif.: Department of Education.

Center on Learning, Assessment, and School Structure. (1993). *Standards, Not Standardization Volumes I-IV: Rethinking Student Assessment (A Video and Print Curriculum in Assessment Reform)*. Geneseo, N.Y.: Center on Learning, Assessment, and School Structure (CLASS).

Elbow, P. (1981). *Writing With Power: Techniques for Mastering the Writing Process*. New York: Oxford University Press.

Fitzpatrick, R., and E. J. Morrison. (1971). "Performance and Product Evaluation." In *Educational Measurement, 2nd Edition. American Council on Education*, edited by R.L. Thorndike. New York: Macmillan Publishing Co. Reprinted in F. L. Finch, ed. (1991). *Educational Performance Assessment*. Chicago, Ill.: Riverside Publishing Company/Houghton Mifflin.

Gentile, C. (April 1991). *Exploring New Methods for Collecting Students' School-Based Writing*. Washington, D.C.: U.S. Department of Education.

Gilbert, T. F. (1978). *Human Competence*. McGraw Hill Publishing Co.

Herman, J., P. Aschbacher, and L. Winters. (1992). *A Practical Guide to Alternative Assessment*. Alexandria, Va.: Association for Supervision and Curriculum Development.

McClelland, D. (1973). "Testing for Competence Rather than for Intelligence." *American Psychologist* 21, 1: 1-14.

Ministry of Education, Victoria, Australia. (1990). *Literacy Profiles Handbook: Assessing and Reporting Literacy Development*. Melbourne, Victoria: Education Shop. (Distributed in the United States by TASA, Brewster, N.Y.)

Mitchell, R. (1992). *Testing for Learning*. New York: Free Press/Macmillan.

Perrone, V., ed. (1991). *Expanding Student Assessment for Supervision and Curriculum Development*. Alexandria, Va.: Association for Supervision and Curriculum Development.

Stiggins, R. (1991). "Assessment Literacy." *Phi Delta Kappan* 72, 1: 534-539.

Wiggins. G. (1993). *Assessing Student Performance: Exploring the Purpose and Limits of Testing*. San Francisco, Calif.: Jossey-Bass Publishers.

Wiggins, G. (May 1992). "Creating Tests Worth Taking." *Educational Leadership* 49, 8: 26-33.

Wiggins, G. (February 1991). "Standards, Not Standardization: Evoking Quality Student Work." *Educational Leadership* 48, 5: 18-25.

Wolf, D., J. Bixby, J. Glen, III, and H. Gardner. (1991). "To Use Their Minds Well, Investigating New Forms of Student Assessment." In *Review of Research in Education*, edited by G. Grant. Washington, D.C.: American Educational Research Association, pp. 31-74.

# VI-6

# Self-Assessment Using Performance Task Assessment Lists

K. MICHAEL HIBBARD
REGION 15 PUBLIC SCHOOLS, MIDDLEBURY AND SOUTHBURY, MIDDLEBURY, CONNECTICUT

Student self-assessment is the heart of performance assessment in the Region 15 Public Schools. Our goal is to coach students to become independent, capable learners who honestly and accurately assess themselves and set meaningful goals for improvement. All of our assessment tools are intended to be used by students as well as teachers and parents.

Region 15 staff members have worked for seven years to define the most important products (such as graphs) and performances (such as oral presentations) which students should be able to make and do in order to demonstrate their growing competence in using content, processes, and work habits in authentic contexts. Assessment tools at different developmental levels are written for each type of student work. Some examples of the products and performances for which assessment tools have been written are in the "menus" for reading comprehension, science, and social studies (see figure 1). Other curriculum areas are also available. The assessment tools are given to teachers in hard copy and on computer disks so that the individual teacher may modify them to fit the needs of the students and/or a task.

The assessment tools, called performance task assessment lists, for graphs are included in this chapter. They are designed for graphs containing an independent variable (i.e., what the experimenter did) and a dependent variable (i.e., how the object reacted). Tools for the primary, elementary, and middle/high school grades are included.

In grades 5 through 12, the lists are used in the following manner. When a performance task calls for a graph, the teacher assigns points to each element on the list to show their relative importance, and gives the list to students. When students finish the graph, they complete the self-assessment column. Teachers make the self-assessment more accurate by showing students examples of excellent graphs from different activities for that grade level. As the teacher assesses each graph, she also assesses the student's self-assessment. When the teacher and student view the quality of the work differently, the teacher coaches the student to view the work more accurately.

The elementary version of the performance task assessment list (grades 3 and 4) has three levels of quality identified for each element scored, plus a holistic rating at the end. For each element, "T"

FIGURE 1

## MENUS FOR THREE COLLECTIONS OF PERFORMANCE TASK ASSESSMENT LISTS

**The three development formats for each list are: primary, elementary, upper grades.**

### SCIENCE

**Scientific Process**

- observations and inferences
- asking questions
- formulating the hypothesis
- designing an experiment
- carrying out the strategy and collecting data
- analyzing the data
- evaluating the hypothesis
- assessing the whole experiment and planning the next experiment

**Graphs and Related Products**

- conducting survey and graphing the results
- data graph
- written summary of a graph

**Science Products**

- lab report
- science fair display
- scientific drawing
- model
- invention
- consumer decision-making study
- making and using a classification system

**Communication Products Using Science Content**

- writer's guide to nonfiction
- writer's guide to fiction
- letter to the editor
- newspaper article
- letter requesting information
- book or pamphlet
- poster
- bulletin board
- cartoon/comic books
- song
- video
- slide show or photo essay
- oral presentation
- investigating an issue controversy
- skit
- display

**Graphic Organizers for Science Content**

- event chain
- concept map
- Venn diagram and pro/con
- idea organizer

**Group Work**

- group work

**Using Math in Science**

- using math in science

**Other**

- information problem solving
- making a performance assessment
- making a performance task
- management plan, individual
- management plan, group
- science journal
- science portfolio

### RESPONDING TO LITERATURE

- initial understanding
- developing an interpretation
- taking a critical stance
- making connections
- locating information: text and graphics
- analyzing information: text and graphics
- making and using: a Venn diagram
- making and using: a plot chain/time line
- making and using: a web
- issue controversy
- writing a summary
- persuasive writing
- fiction summary
- compare/contrast
- theme
- expository writing
- narrative writing

### SOCIAL STUDIES

- analysis of a leader
- autobiography
- book with illustrations
- bulletin board display
- business letter
- cause/effect essay
- conducting an interview
- display/diorama
- drawing/illustration
- editorial/letter to the editor
- event chain graphic organizer
- expository essay
- folktale
- geography
- grant application simulation
- graph, bar/line
- graph, pie
- group work
- hypermedia presentation
- idea/information organizer, graphic organizer
- information problem solving
- interview
- invention
- investigating an issue, controversy
- journal
- land use survey
- map
- math problem solving
- model
- newspaper article
- oral presentation
- pamphlet
- peer editing
- persuasive writing
- political cartoon
- portfolio
- poster
- research report
- scrapbook/artifact collection
- skit/play
- slide show/photo essay
- survey and graph
- task management plan, individual
- task management plan, group
- time line, illustrated
- travel brochure for a country
- tribute to a person
- weather map
- Venn diagram/pro-con graphic organizer
- video
- writer's guide to fiction
- writer's guide to nonfiction

means "terrific," "O" means "okay," and "W" means "needs work." As with the secondary version, students receive this list up front, use it to guide their work, and then do their self-assessment. Again, teachers assess the graph and the students' self-assessment. The primary version of the list (grades 1 and 2 is simpler and does not ask the students for an overall judgment. When these lists are begun in the 1st grade, teachers show students one element at a time and build up to the entire list by the end of the year. Models of excellent graphs for each grade level are important components of this assessment system.

In grades 5 through 12, lists are used with "rubrics." The list allows students to understand and assess the elements of quality for each separate task, and the rubric helps them develop a holistic view of overall work in graphing. After several months of making graphs in one or more courses, students are given the rubric and asked to evaluate how their graphs relate, overall, to the standards of quality for that grade level. In the attached graphing rubric, "S" means superb, incredible, outstanding; "T" means evenly excellent; "U" means that the work is mostly excellent, but there is an important element which is not excellent; "V" means that the work is mostly poor, but there is an important element which is good; "W" means evenly poor; and "X" means the work is very poor. The first "decision" is whether the graph is more like a T or more like a W according to the standards for that grade level. If it is more like a T, then the second decision is whether it remains a T, is really a U, or is so outstanding it should be raised to an S. If the initial decision is that it is more like a W, then the last decision is whether it is a V, a W, or an X.

The goal is to coach students to become independent learners. Our assessment tools help students learn about the elements of quality of many products and performances. Although we have created hundreds of these assessment tools, our goal is to wean students from them well before high school graduation. The weaning happens gradually. First, teachers give the lists to the students. Second, students make their own lists, using them just as they did the teacher-supplied lists. Creating these lists is a wonderful cooperative learning activity. Third, teachers asks students to make notes for themselves about the elements of quality they should attend to as they do their work. These notes are for the student only. Finally, the students engage in work without any teacher direction.

If you walk into a classroom where students are using assessment lists and models of excellence to shape and assess their work and you ask whether you're watching the teaching activity or the assessment activity, the answer is "Absolutely!"

**Performance Criteria for Graphing, Grades 5-12**

## CLASSROOM ASSESSMENT LIST
## DATA TABLE, HIGH SCHOOL

| | ELEMENT | ASSESSMENT POINTS | | |
|---|---|---|---|---|
| | | Points Possible | Earned Assessment: | |
| | | | Self | Teacher's |
| 1. | The data table carries the appropriate date. | ---- | --- | --- |
| 2. | An appropriate title for the data table is provided. | ---- | --- | --- |
| 3. | The information on the data table columns is appropriately organized and labeled. | ---- | --- | --- |
| 4. | Units of measurement for all variables are clearly indicated. | ---- | --- | --- |
| 5. | Data for the independent and dependent variables are clearly shown. | ---- | --- | --- |
| 6. | The data have an appropriate number of significant figures. | ---- | --- | --- |
| 7. | Accuracy of the data is appropriate to the measuring equipment or instrument being used. | ---- | --- | --- |
| 8. | Data from multiple trials at each level of the independent variable are clearly shown. | ---- | --- | --- |
| 9. | The data table is neat and presentable. | ---- | --- | --- |
| | Total | ---- | --- | --- |

*Pomperaug Regional School District 15*
*Middlebury, Connecticut*
*K. Michael Hibbard*

# CLASSROOM ASSESSMENT LIST FOR A GRAPH, HIGH SCHOOL

| | Element | ASSESSMENT POINTS | | |
|---|---|---|---|---|
| | | Points Possible | Earned Assessment: | |
| | | | Self | Teacher's |
| 1. | An appropriate type of graph is used. | ___ | ___ | ___ |
| 2. | Appropriate starting points and intervals are used for each axis. | ___ | ___ | ___ |
| 3. | There is a main title for the graph which clearly states the relationship between the axes. | ___ | ___ | ___ |
| 4. | An appropriate scale is used on each axis depending on the range of data for that axis. | ___ | ___ | ___ |
| 5. | Axes are clearly labeled. | ___ | ___ | ___ |
| 6. | The independent variable is put on the(X) axis and the dependent variable is put on the (Y) axis. | ___ | ___ | ___ |
| 7. | The data are plotted accurately. | ___ | ___ | ___ |
| 8. | The graph should reflect uncertainty of measurement. | ___ | ___ | ___ |
| 9. | Trends or lack of trends are depicted on the graph. | ___ | ___ | ___ |
| 10. | An appropriate key or legend is part of the graph. | ___ | ___ | ___ |
| 11. | The lines or bars use the space of the graph well. | ___ | ___ | ___ |
| 12. | Appropriate techniques such as color, texture, or clarifying labels are used to make the graph easier to understand. | ___ | ___ | ___ |
| 13. | The whole graph uses the space given it on the paper well. | ___ | ___ | ___ |
| 14. | The graph is neat and presentable. | ___ | ___ | ___ |
| 15. | The graph is easy to interpret. | ___ | ___ | ___ |
| | Total Points | ___ | ___ | ___ |

BEST COPY AVAILABLE

# CLASSROOM ASSESSMENT LIST FOR A WRITTEN SUMMARY OF A GRAPH
## HIGH SCHOOL

| ELEMENT | ASSESSMENT POINTS | | |
|---|---|---|---|
| | Points Possible | Earned Assessment: | |
| | | Self | Teacher's |
| 1 The summary gives a reasonable interpretation of the data. | ____ | ___ | ___ |
| 2. The research question is answered. | ____ | ___ | ___ |
| 3 There is reference to both independent and dependent variables. | ____ | ___ | ___ |
| 4. The relationship between the dependent and the independent variables is clearly and accurately described. | ____ | ___ | ___ |
| 5. The summary is concise. | ____ | ___ | ___ |
| 6 Thoughtful mastery of language is evident including the use of appropriate vocabulary, language mechanics, and complete sentences. | ____ | ___ | ___ |
| 7 The summary is neat and presentable | ____ | ___ | ___ |
| Total | ____ | ___ | ___ |

BEST COPY AVAILABLE

## RUBRIC FOR A DATA TABLE

S: The data table is exceptionally well organized, complete, and clear.

T: The data table has a concise and accurate title. Data about the independent variable (what the experimenter does) and data about the dependent variable (how the thing responds) are clearly marked and units for all data are clearly indicated. An appropriate number of significant figures is used. The accuracy of the data is appropriate to the measuring equipment or instrument being used. The table is organized in such as way that data from multiple trials and measures of central tendency are evident. The data table is neat and presentable.

U: The data table is similar to one receiving the rating of T, except an important element is excellent.

V: The data table is similar to one receiving the rating of W, except an important element is well done.

W: The data table lacks a title or the title is inaccurate or unclear. The data about the independent and dependent variables are not well organized and the units are missing or inaccurate. An inappropriate number of significant figures is used. It is unclear how the strategy of multiple trials was used. An inappropriate or inaccurate measure of central tendency was used. Over-all, the data table is unorganized and not very neat.

X: The data table is very poor or missing.

## RUBRIC FOR A GRAPH

**S:** The student's graph is outstanding in its ability to clearly and easily convey accurate information to the reader. The graph is very interesting.

**T:** The student has selected the appropriate type of graph for that data. The variables are put on the correct axes which are accurately labeled. An appropriate scale with reasonable starting points and intervals is used on each axis so that the graph uses the space allotted it well. The data is plotted accurately and trend(s), or lack of it (them), is (are) accurately indicated. The title is clear and describes the two variables well. Techniques such as color, texture, or labels are used so that graph can be read at a glance. A clear key is presented. The graph is very neat and presentable.

**U:** The graph is like one receiving the rating of T, except there are one or two important elements which are not excellent.

**V:** The graph is like one receiving the rating of W, except there are one or two important elements which are well done.

**W:** An inappropriate or incorrect type of graph was chosen. The variables are put on the wrong axes. The axes are not labeled or labeled inadequately. The scales used are not appropriate and the lines or bars do not use the space of the graph well. The title is missing or inadequate. Techniques such as color, texture, or labels are not used well to enhance the ease of interpretation of the graph. The key is missing or inadequate.

**X:** The graph is very poorly done.

## RUBRIC FOR A WRITTEN SUMMARY OF A GRAPH

S: The summary is accurate and precise and uses especially appropriate vocabulary to describe the graph to the intended audience.

T: The summary clearly and accurately states the relationship between the independent variable (what I do) and dependent variable (how it responds.) A conclusion is reached about the research question. The summary states what the data shows and, if extrapolations are made, they are identified as such. The summary is concise and neat. Language mechanics are excellent.

U: The summary is similar to one receiving a rating of T, except that some important element is not excellent.

V: The summary is similar to one receiving a rating of W, except that some important element is well done.

W: The summary is inaccurate or unclear in stating the relationship between the independent and dependent variables. No conclusion about the research question is clearly reached. The summary mixes statement of facts with extrapolations. It is not well organized and its language mechanics are not well done.

X: The summary is very poor or missing.

## CLASSROOM ASSESSMENT LIST
## DATA TABLE, MIDDLE SCHOOL

| | ELEMENT | ASSESSMENT POINTS<br>Points Possible | Earned Assessment: Self | Teacher's |
|---|---|---|---|---|
| 1 | The data table carries the appropriate date. | ---- | --- | --- |
| 2. | An appropriate title for the data table is provided. | ---- | --- | --- |
| 3 | The information on the data table columns is appropriately organized and labeled | ---- | --- | --- |
| 4. | Units of measurement for all variables are clearly indicated. | ---- | --- | --- |
| 5. | Data for the independent variable (what you change) are shown clearly | ---- | --- | --- |
| 6. | Data for the dependent variable (how the thing responds) are shown clearly. | ---- | --- | --- |
| 7 | The data have an appropriate number of significant figures. | ---- | --- | --- |
| 8. | Accuracy of the data is appropriate to the measuring equipment or instrument being used. | ---- | --- | --- |
| 9 | Data from multiple trials at each level of the independent variable are clearly shown. | ---- | --- | --- |
| 10 | The data table is neat and presentable. | ---- | --- | --- |
| | Total | ---- | --- | --- |

## CLASSROOM ASSESSMENT LIST FOR A GRAPH, MIDDLE SCHOOL

| | Element | ASSESSMENT Points Possible | POINTS Earned Assessment: Self | Teacher's |
|---|---|---|---|---|
| 1. | An appropriate type of graph is used. | ___ | ___ | ___ |
| 2. | Appropriate starting points and intervals are used for each axis. | ___ | ___ | ___ |
| 3. | There is a main title for the graph which clearly states the relationship between the axes. | ___ | ___ | ___ |
| 4. | An appropriate scale is used on each axis depending on the range of data for that axis. | ___ | ___ | ___ |
| 5. | Axes are clearly labeled. | ___ | ___ | ___ |
| 6. | The independent variable (what you change) is put on the (X) horizontal axis and the dependent variable (how it responds) is put on the (Y) vertical axis. | ___ | ___ | ___ |
| 7. | The data are plotted accurately. | ___ | ___ | ___ |
| 8. | The graph should reflect uncertainty of measurement. | ___ | ___ | ___ |
| 9. | Trends or lack of trends are depicted on the graph. | ___ | ___ | ___ |
| 10. | An appropriate key or legend is part of the graph. | ___ | ___ | ___ |
| 11. | The lines or bars use the space of the graph well. | ___ | ___ | ___ |
| 12. | Appropriate techniques such as color, texture, or clarifying labels are used to make the graph easier to understand. | ___ | ___ | ___ |
| 13 | The whole graph uses the space given it on the paper well. | ___ | ___ | ___ |
| 14. | The graph is neat and presentable. | ___ | ___ | ___ |
| 15. | The graph is easy to interpret. | ___ | ___ | ___ |
| | Total Points | ___ | ___ | ___ |

BEST COPY AVAILABLE

# CLASSROOM ASSESSMENT LIST FOR A WRITTEN SUMMARY OF A GRAPH
## MIDDLE SCHOOL

| ELEMENT | | ASSESSMENT POINTS | | |
|---|---|---|---|---|
| | | Points Possible | Earned Assessment: Self | Teacher's |
| 1. | The summary gives a reasonable interpretation of the data. | ____ | ___ | ___ |
| 2. | A conclusion is reached. | ____ | ___ | ___ |
| 3. | There is reference to the independent variable (what I do) and to the dependent variable (how it responds.) | ____ | ___ | ___ |
| 4. | The relationship between the independent variable (what I do) and the dependent variable (how it responds) is clearly and accurately described. | ____ | ___ | ___ |
| 5. | The summary is to the point. | ____ | ___ | ___ |
| 6. | Thoughtful mastery of language is evident including the use of appropriate vocabulary, language mechanics, and complete sentences. | ____ | ___ | ___ |
| 7 | The summary is neat and presentable. | ____ | ___ | ___ |
| | Total | ____ | ___ | ___ |

## CLASSROOM ASSESSMENT LIST
## DATA TABLE, ELEMENTARY SCHOOL

1. Heading
   T: I put my name and the date on my work.
   O: I left my name or the date off.
   W: I did not include my name or date.

2. Title
   T: My title says exactly what the data is about.
   O: My title gives some idea of what the data is about.
   W: I have no title or my title does not tell much about what the data is about.

3. Organizing The Data
   T: My data is organized into very neat rows and columns.
   O: Some of my data is organized.
   W: I need to organize my data much better.

4. Naming The Data
   T: The rows and columns have titles and all the data has units.
   O: Some rows and columns have titles and some data has units.
   W: I need to give the rows and columns titles and/or I need to give the data units.

Did I do my best work?

| Terrific | OK | Needs Work |
|---|---|---|
| 😀 | 🙂 | 😐 |

## CLASSROOM ASSESSMENT LIST FOR A GRAPH, ELEMENTARY SCHOOL

1. Heading
   T: I included my name and the date.
   W: I left off my name or the date.
   O: I did not include my name or the date.

2. Title
   T: The title tells exactly what the graph is about. The title includes a short statement of what I changed and what I measured
   O: The title tells what the graph is about.
   W: The title is missing or it tells little.

3. Labeling The Data
   T: Each axis has a name that explains what that axis is and the data has units.
   O: The axis need names that are more clear and/or the data needs units.
   W: The names for the axes and/or the units for the data are missing.

4. Selecting A Scale For Each Axis
   T: Each axis has a scale that fits the data in the data table. The scales help use the whole graph well.
   O: The scales for one or both axes need to be improved.
   W: The scales are missing or they are incorrect.

5. Drawing Lines Or Bars On The Graph
   T: The lines or bars are drawn accurately and very neatly.
   O: There are some mistakes or the work is a little messy.
   W: There are many mistakes and/or the work is very messy.

5. Use Of Space And Color
   T: The whole graph uses the space well. Color or some other technique is used so the graph is easy to read.
   O: The space is not used too well and color or some other technique could be used better so that the graph is easier to read.
   W: The graph needs a lot of work on the use of space and color.

6. Key Or Legend
   T: The key or legend is very clear.
   O: The key or legend is OK.
   W: The key or legend needs a lot or work.

Did I do my best?

| Terrific | OK | Needs Work |
|---|---|---|
| | | |

## CLASSROOM ASSESSMENT LIST
## WRITTEN SUMMARY OF A GRAPH
## ELEMENTARY

1. Heading
   T: My name and the date are included on my work.
   O: Either my name or the date is left off.
   W: My name and the date are left off.

2: Title
   T: The title of my summary tells the reader what the summary is about.
   O: The title tells the reader a little.
   W: The title needs a lot of work.

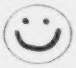
3. Summary
   T: My summary states exactly what the graph shows.
   O: My summary states some of what the graph shows.
   W: My summary states only a little of what the graph shows.

4. Language Mechanics
   T: My spelling, capitalization, and punctuation are excellent.
   O: My spelling, capitalization, and punctuation need some work.
   W: My spelling, capitalization, and/or punctuation need a lot of work.

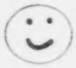
5. Sentences And Vocabulary
   T: My sentences are complete and interesting. I have used appropriate vocabulary words.
   O: My sentences are complete but they could be more interesting. I could use better vocabulary.
   W: My sentences and vocabulary need a lot of work.

Did I do my best work?

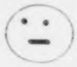
Terrific



OK

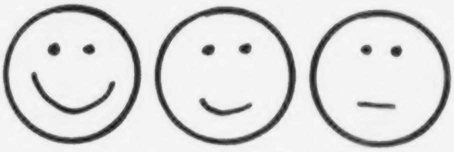


Needs Work

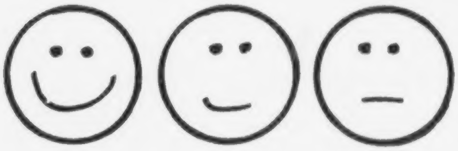


## CLASSROOM ASSESSMENT LIST
## DATA TABLE, PRIMARY GRADES

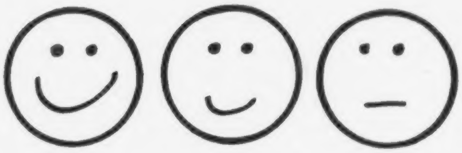
1. Did I draw what I counted?



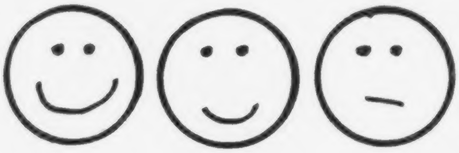
2. Did I organize my drawing?



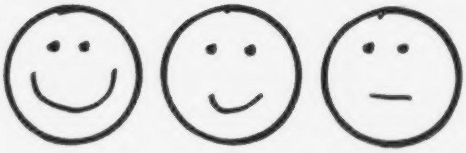
3. Did I print words next to my drawings?



4. Did I say what I drew?

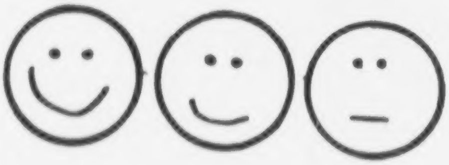


5. Is my work neat?

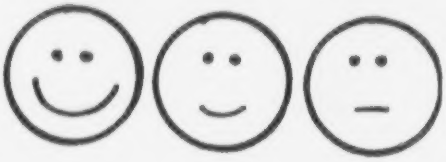


## CLASSROOM ASSESSMENT LIST FOR A GRAPH, PRIMARY GRADES

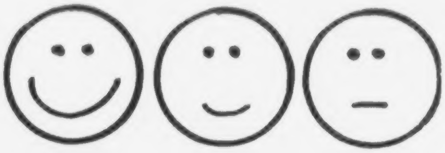
1. Did I draw the things in columns?



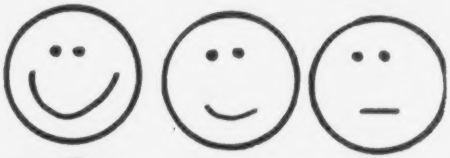
2. Did I make a separate column for each different kind of thing?



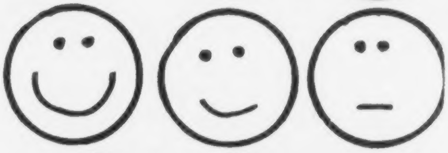
3. Did I print words to tell what each column is?



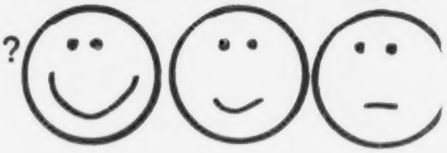
4. Did I use the space of the paper well?



6. Did I use color?

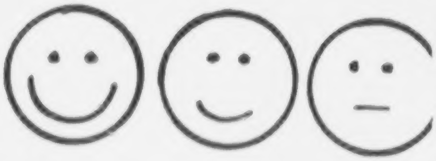


7. Is my graph drawn neatly?

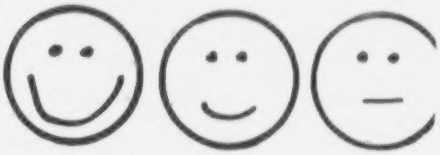


# CLASSROOM ASSESSMENT LIST
# ORAL SUMMARY OF A GRAPH
# PRIMARY GRADES

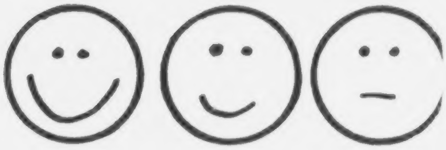
1. Did I say what the graph shows?



2. Did I talk in complete sentences?



3. Did I point to the graph to show what I was talking about?



# VI-7

# Writing Assessment

VICKI SPANDEL AND RUTH CULHAM
NORTHWEST REGIONAL EDUCATIONAL LABORATORY

The attached performance criteria for writing are collectively called a "six-trait analytical model," because writing has been analyzed to have six traits or dimensions worth scoring separately. The scoring guide has been used with various types of writing (persuasive, expository, and narrative), in both classrooms and large-scale assessment, and for both prompted and student-generated writing.

The scoring guide was first developed in 1985 in Beaverton School District (Oregon) by teachers who wanted a better way to assess student writing than the holistic scoring procedure they were using in which writing received one overall score. At the current time, this model is used in several states (e.g., Kansas, Oregon, Washington, Alaska) as the major tool for writing assessment. It is based on analysis of thousands of samples of student writing in all grade levels to determine what "makes it work." Much of this on-going work has been done by the individuals listed below. They have also designed simplified scoring guides for student use at various grade levels.

Vicki Spandel and Ruth Culham
Northwest Regional Educational Laboratory
101 S.W. Main, Suite 500
Portland, Oregon 97204
Phone: (503) 275-9500

# IDEAS AND CONTENT

## (Development)

⑤ ***This paper is clear and focused. It holds the reader's attention. Relevant anecdotes and details enrich the central theme or storyline.***

- o Ideas are **fresh** and **original**.
- o The writer seems to be writing from **knowledge** or **experience** and shows **insight**: an understanding of life and a knack for picking out what is significant.
- o **Relevant, telling, quality details** give the reader important information that goes **beyond the obvious** or predictable.
- o The writer develops the topic in an **enlightening, purposeful** way that **makes a point** or **tells a story**.
- o Every piece adds something to the whole.

③ ***The writer is beginning to define the topic, even though development is still basic or general.***

- o It is pretty **easy to see where the writer is headed**, though more information is needed to "fill in the blanks."
- o The writer seems to be drawing on knowledge or experience, but has **difficulty going from general observations to specifics.**
- o Ideas are **reasonably clear**, though they may not be detailed, personalized, or expanded enough to show indepth understanding or a strong sense of purpose.
- o **Support is attempted**, but doesn't go far enough yet in fleshing out the main point or storyline.
- o Details often **blend the original with the predictable**.

① ***As yet, the paper has no clear sense of purpose or central theme. To extract meaning from the text, the reader must make inferences based on sketchy details. The writing reflects more than one of these problems:***

- o The writer is **still in search of a topic**, or has not begun to define the topic in a meaningful, personal way.
- o Information is very **limited** or **unclear.**
- o The text may be **repetitious**, or may read like a collection of **disconnected, random thoughts.**
- o **Everything seems as important as everything else**; the reader has a hard time sifting out what's critical.

# ORGANIZATION

⑤ ***The organization enhances and showcases the central idea or storyline. The order, structure or presentation of information is compelling and moves the reader through the text.***

- Details seem to fit where they're placed; **sequencing is logical** and **effective**.
- An **inviting introduction** draws the reader in; a **satisfying conclusion** leaves the reader with a sense of resolution.
- **Pacing is well controlled**; the writer knows when to slow down and elaborate, and when to pick up the pace and move on.
- **Thoughtful transitions** clearly show how ideas connect.
- Organization **flows so smoothly** the reader hardly thinks about it.

③ ***The organizational structure is strong enough to move the reader through the text without undue confusion.***

- The paper has a **recognizable introduction and conclusion**. The introduction may not create a strong sense of anticipation; the conclusion may not tie up all loose ends.
- Sequencing is usually logical, but may sometimes be so predictable that the **structure takes attention away from the content.**
- **Pacing is fairly well controlled,** though the writer sometimes spurts ahead too quickly or spends too much time on details that do not matter.
- **Transitions often work well**; at other times, connections between ideas are fuzzy.
- The **organization sometimes supports the main point or storyline;** at other times, the reader feels an urge to slip in a transition or move things around.

① ***The writing lacks a clear sense of direction. Ideas, details or events seem strung together in a loose or random fashion--or else there is no identifiable internal structure. The writing reflects more than one of these problems:***

- **Sequencing needs work.**
- There is **no real lead** to set up what follows, **no real conclusion** to wrap things up.
- **Pacing feels awkward;** the writer slows to a crawl when the reader wants to get on with it, and vice versa.
- **Connections** between ideas are **confusing** or missing.
- Problems with organization make it **hard for the reader to get a grip** on the main point or storyline.

# VOICE

⑤ ***The writer speaks directly to the reader in a way that is individualistic, expressive and engaging. Clearly, the writer is involved in the text, is sensitive to the needs of an audience, and is writing to be read.***

- The reader feels a **strong interaction** with the writer, sensing the **person behind the words.**
- The tone and voice give **flavor** to the message and seem **appropriate for the purpose and audience.**
- **Narrative** writing seems **honest, appealing,** and written **from the heart**.
- **Expository or persuasive** writing reflects a **strong commitment** to the topic, and an effort to bring the topic to life by **anticipating the reader's questions,** and **showing why the reader should care** or want to know more.

③ ***The writer seems sincere, but not fully engaged or involved. The result is pleasant or even personable, but not compelling.***

- The writing communicates in an **earnest, pleasing** manner. Moments here and there surprise, amuse or move the reader.
- Voice may emerge strongly on occasion, then retreat behind **general, dispassionate language.**
- The writing **hides** as much of **the writer** as it reveals.
- The writer seems **aware of an audience,** but often to **weigh words carefully** or discard personal insights in favor of **safe generalities**.

① ***The writer seems indifferent, uninvolved or distanced from the topic and/or the audience. As a result, the writing is lifeless or mechanical; depending on the topic, it may be overly technical or jargonistic. The paper reflects more than one of the following problems:***

- It is **hard to sense the writer** behind the words. The writer **does not seem to reach out** to an audience, or to anticipate their interests and questions.
- The writer speaks in a kind of **monotone** that flattens all potential highs or lows of the message.
- The writing may communicate on a **functional** level, but it **does not move** or involve the reader.
- The **writer does not seem** sufficiently **at home with the topic** to take risks, share personal insights or make the topic/story personal and real for the reader.

# WORD CHOICE

⑤ ***Words convey the intended message in a precise, interesting and natural way.***

- Words are **specific** and **accurate**; it is easy to understand just what the writer means.
- The language is **natural and never overdone**; phrasing is highly **individual**.
- **Lively verbs** energize the writing. **Precise nouns and modifiers** create pictures in the reader's mind.
- **Striking words and phrases** often catch the reader's eye--and linger in the reader's mind.
- **Clichés** and **jargon** are **used sparingly**, only for effect.

③ ***The language is functional, even if it lacks punch; it is easy to figure out the writer's meaning on a general level.***

- Words are almost always **correct and adequate;** they simply **lack flair.**
- **Familiar words and phrases** communicate, but rarely capture the reader's imagination. Still, the paper may have one or two fine moments.
- **Attempts at colorful language** come close to the mark, but sometimes seem **overdone**.
- Energetic verbs or picturesque phrases **liven things up now and then**; the reader longs for more.

① ***The writer struggles with a limited vocabulary, searching for words to convey meaning. The writing reflects more than one of these problems:***

- Language is so **vague** (e.g., *It was a fun time, She was neat, It was nice, We did lots of stuff*) that only the **most general message** comes through.
- Persistent **redundancy** distracts the reader.
- **Jargon or clichés** serve as a crutch.
- **Words are used incorrectly,** sometimes making the message hard to decipher.
- Problems with language **leave the reader wondering** what the writer is trying to say.

# SENTENCE FLUENCY

⑤ ***The writing has an easy flow and rhythm when read aloud. Sentences are well built, with strong and varied structure that invites expressive oral reading.***

- o Sentences are constructed in a way that helps make **meaning clear**.
- o **Purposeful sentence beginnings** show how each sentence relates to and builds upon the one before it.
- o The writing has **cadence,** as if the writer has thought about the sound of the words as well as the meaning.
- o Sentences **vary in length as well as structure.**
- o Fragments, if used, add style.
- o Dialogue, if used, sounds natural.

③ ***The text hums along with a steady beat, but tends to be more pleasant or businesslike than musical, more mechanical than fluid.***

- o Sentences may not seem artfully crafted or musical, but they are usually **grammatical.** They **hang together**. They get the job done.
- o There is at least **some variation in sentence length and structure**. Sentence beginnings are NOT all alike.
- o The reader sometimes has to **hunt for clues** (e.g., connecting words and phrases like *however, therefore, naturally, after a while, on the other hand, to be specific, for example, next, first of all, later, but as it turned out, although,* etc.) that show how sentences interrelate.
- o **Parts of the text invite expressive oral reading**; others may be stiff, awkward, choppy or gangly. Overall though, it is pretty easy to read the paper aloud with a little practice.

① ***The reader has to practice quite a bit in order to give this paper a fair interpretive reading. The writing reflects more than one of the following problems:***

- o Sentences are **choppy, incomplete, rambling or awkward**; they need work.
- o **Phrasing does not sound natural**, the way someone might speak. The reader must sometimes **pause or read over** to get the meaning.
- o Many **sentences begin the same way**--and may follow the same patterns (e.g., *subject-verb-object*) in a monotonous pattern.
- o **Endless connectives** (*and, and so, but then, because, and then,* etc.) create a massive jumble of language in which clear sentence beginnings and endings get swallowed up.
- o The text **does not invite expressive oral reading**.

# CONVENTIONS

⑤ ***The writer demonstrates a good grasp of standard writing conventions (e.g., grammar, capitalization, punctuation, usage, spelling, paragraphing) and uses conventions effectively to enhance readability. Errors tend to be so few and so minor that the reader can easily overlook them unless hunting for them specifically.***

- **Paragraphing tends to be sound** and to reinforce the organizational structure
- **Grammar and usage are correct** and contribute to clarity and style
- **Punctuation is accurate** and guides the reader through the text
- **Spelling is generally correct**, even on more difficult words
- The writer **may manipulate conventions**--especially grammar and spelling--for stylistic effect.
- **GRADES 7 AND UP ONLY:** The writing is **sufficiently long and complex** to allow the writer to show skill in using a wide range of conventions
- Only **light editing** would be required to polish the text for publication

③ ***The writer shows reasonable control over a limited range of standard writing conventions. Conventions are sometimes handled well and enhance readability; at other times, errors are distracting and impair readability.***

- **Paragraphing is attempted.** Paragraphs sometimes run together or begin in the wrong places.
- **Problems with grammar or usage are not serious** enough to distort meaning
- **Terminal (end-of sentence) punctuation is usually correct**; internal punctuation (commas, apostrophes, semicolons, dashes, colons, parentheses) is sometimes missing or wrong.
- **Spelling** is usually **correct or reasonably phonetic on common words.**
- **Moderate editing** would be required to polish the text for publication

① ***Errors in spelling, punctuation, usage and grammar, capitalization, and/or paragraphing repeatedly distract the reader and make the text difficult to read. The writing reflects more than one of these problems:***

- **Paragraphing is missing, irregular, or so frequent** (e.g., every sentence) that it has no relationship to the organizational structure of the text.
- **Errors in grammar or usage are very noticeable,** and may affect **Extensive editing** would be required to polish the text for publication.meaning
- **Punctuation** (including terminal punctuation) is often **missing or incorrect**
- **Spelling errors are frequent,** even on common words.
- The reader must **read once to decode**, then again for meaning
- **Extensive editing** would be required to polish the text for publication

# VI-8

# Performance Assessment in Science

GREG HALL
ALBERTA EDUCATION, CANADA

This set of performance criteria is associated with a set of tasks designed to measure problem solving in science at grade 9. Students circulate to six stations that contain hands-on materials in the areas of: seed dispersal (diversity of living things), sugar solution (fluid and pressure), sleeping bags (heat energy), robot arm (electromagnetic systems, bulk food store (chemical properties), and hazardous waste (environmental quality). It takes about two hours for students to complete all the tasks in the circuit. The same performance criteria are used for all performance tasks.

Greg Hall
Alberta Education
Box 43, 11160 Jasper Avenue
Edmonton, Alberta, Canada T5K 0L2
Phone: (403) 427-0010
Fax: (403) 422-4200

---

Reprinted with permission of Alberta Education, Devonian Building, West Tower, 11160 Jasper Avenue, Edmonton, Alberta, Canada TK5 OL2. Phone: (403) 427-0010.

## Science Holistic Scoring Criteria

| | | Problem Solving/Inquiry | Communication |
|---|---|---|---|
| 3 | **BEYOND** | — Analyzed and readily understood the task<br>— Developed an efficient and workable strategy<br>— Strategy implemented effectively<br>— Strategy supports a qualified solution<br>— Appropriate application of critical knowledge | — Appropriate, organized, and effective system for display of information or data<br>— Display of information or data is precise, accurate, and complete<br>— Interpretations and explanations logical and communicated effectively |
| 2 | **AT GRADE** | — Understood the task<br>— Developed a workable strategy<br>— Strategy inferred (some evidence) but not always clear<br>— Strategy supports appropriate solution<br>— Evidence of application of critical knowledge | — Appropriate, organized system for display of information or data<br>— Display of information or data is mostly precise, accurate, and complete<br>— Interpretations and explanations logical and mostly clear |
| 1 | **NOT YET AT** | — Partially understood the task<br>— Appropriate strategy some of the time<br>— Possible evidence of a plan, but not clear<br>— Partial connection to appropriate solution<br>— Partial evidence of application of critical knowledge | — System for display of information or data may not be clear or effective<br>— Display of information or data is somewhat precise, accurate, and complete<br>— Interpretations and explanations somewhat clear |
| 0 | | — Misunderstood the task<br>— Inappropriate, unworkable strategy<br>— No evidence of carrying out a plan<br>— No connections to solution<br>— No evidence of critical knowledge<br>— Blank | — Disorganized system for display of information or data<br>— Display of information or data is not precise, accurate, or complete<br>— Interpretations and explanations not clear<br>— Blank |

BEST COPY AVAILABLE

## VI-9

# Assessment of Laboratory Skills in High School Science

Rodney Doran
SUNY at Buffalo

The author has developed laboratory performance assessments in chemistry, physics, and biology. Six tasks are presented in each content area. Each task has two parts which take a total of 80 minutes. In Part A, students are given a problem to solve and are directed to state an appropriate hypothesis, develop a procedure for gathering relevant observations or data and propose a method for organizing the information collected. These plans are collected prior to beginning Part B.

In Part B, students are given predeveloped procedures to collect information on the same problem as in Part A. This procedure ensures that success on Part B is not dependent on Part A. In Part B student work is examined for quality of observations or data, performance of appropriate calculations, and formulation of valid and consistent conclusions from experimental results.

The same criteria are used for all tasks.

These tasks were constructed as part of a project supported by the National Science Foundation in collaboration with Dr. Darrell Bock of the University of Chicago.

Dr. Rodney Doran
Graduate School of Education
State University Of New York at Buffalo
Buffalo, NY 14260
Phone: (716) 645-3171

---

Reprinted with permission from the University at Buffalo, Graduate School of Education, 593 Baldy Hall, Buffalo, New York 14260-1000. Phone: (716) 645-2455.

**Figure 1: Scoring Form for Science Laboratory Test**

School/Student ID No. ________ Reader ID No. ________ Date ________ Time ________

Subject Area B C P Task 1 2 3 4 5 6

1. Please circle the NA code if a skill is not assessed in a particular area.
2. The NR code is to be circled when no attempt to respond to the question is apparent.
3. You may check each element present (✓) and sum up to determine a student's score for each skill.
4. There is no need to determine a total score for a student.

**Part A: Experiment Design**

1. STATEMENT OF HYPOTHESIS NR 0 1 2 3 4 5 NA
   - Effect linked to variable ____
   - Directionality of effect ____
   - Expected effect/change ____
   - Independent variable ____
   - Dependent variable ____

2. PROCEDURE FOR INVESTIGATION NR 0 1 2 3 4 5 NA
   - Resolves experimental problem/feasible ____
   - Sequenced and detailed plan ____
   - General strategy ____
   - Safety procedures ____
   - Use of equipment/diagram of set-up ____

3. PLAN TO RECORD AND ORGANIZE OBSERVATIONS/DATA NR 0 1 2 3 4 5 NA
   - Space for measured/calculated data ____
   - Matched to plan ____
   - Organized sequentailly ____
   - Labelled fully (units included) ____
   - Variables identified ____

**Part B: Experiment Report**

4. QUALITY OF OBSERVATIONS/DATA NR 0 1 2 3 4 5 NA
   - Consistent data ____
   - Accurate measurements/observations ____
   - Completed data table ____
   - Correct units ____
   - Qualitative description ____

5. GRAPH NR 0 1 2 3 4 5 NA
   - Curve is appropriate to data trend ____
   - Points plotted accurately ____
   - Appropriate scale (units included) ____
   - Axes labelled with correct variables ____
   - Has an appropriate title ____

6. CALCULATIONS NR 0 1 2 3 4 5 NA
   - Calculated accurately ____
   - Substituted correctly into relationship ____
   - Relationship stated or implied ____
   - Units used correctly ____
   - Used all data available ____

7. FORMS A CONCLUSION FROM THE EXPERIMENT NR 0 1 2 3 4 5 NA
   - Consistent with scientific principle ____
   - Sources of error ____
   - Consistent with data ____
   - Relationship among variables stated ____
   - Variables stated in conclusion ____

# VI-10

## Primary Developmental Continuums—Reading, Speaking, Writing

ANNIE CALKINS
JUNEAU SCHOOL DISTRICT, ALASKA

Juneau School District is in its third year of development of portfolios to assess primary students' development in language arts. The portfolio is intended to replace report cards and standardized tests. Samples of student work or performance are placed into the portfolio at regular times during the school year. For example, student reading samples are included at least twice—during the first and fourth quarters. An "x" is placed on the developmental continuum in the place that best reflects the student's current level of performance. For example, at the beginning of the year, the student might be an "Emergent" reader, while at the end of the school year the student might be a "Developing" reader. The reading samples in the portfolio illustrate and substantiate the rating.

Complete information is available in the publication *Language Arts Portfolio Handbook for the Primary Grades, 3rd Edition.*

Annie Calkins
Juneau School District
10014 Crazy Horse Drive
Juneau, Alaska 99801
Phone: (907) 463-1967
Fax: (907) 789-6977

---

Reprinted with permission from the Juneau School District, 10014 Crazy Horse Drive, Juneau, Alaska 99801. Phone: (907) 463-1967.

# JUNEAU PRIMARY READING CONTINUUM

Name ____________________

| | Emergent | Beginning | Developing | Expanding | Independent |
|---|---|---|---|---|---|
| **Comprehension** | • **Relies on memory for reading**<br>• Responds to stories<br>• May label pictures<br>• May tell a story from pictures using oral language<br>• May pretend to read<br>• May invent text with book language<br>• Focuses on pictures for meaning rather than print | • **Reads simple books in which text is repeated; (illustrations provide a lot of support) and demonstrates understanding in the following ways:**<br>- Recalls random details<br>- Recognizes when the reading isn't making sense<br>- Shows understanding that print carries meaning | • **Reads books with varied sentence patterns; (illustrations provide a moderate amount of support) and demonstrates understanding in a few of the following ways:**<br>- Recounts sequence of events<br>- Summarizes story<br>- Predicts what will happen next<br>- Backs up statements with proof from reading<br>- Connects experiences with reading | • **Reads books with long descriptions, challenging vocabulary; (illustrations provide low support) and demonstrates understanding in several of the following ways:**<br>- Remembers sequence of events<br>- Summarizes story<br>- Predicts what will happen next<br>- Backs up statements with proof from reading<br>- Connects and builds to draw conclusions<br>- Uses prior knowledge with relevent information from the story to form an opinion<br>- Connects experiences with reading | • **Reads books with long descriptions, challenging vocabulary; (illustrations provide very little or no support) and demonstrates understanding in most of the following ways:**<br>- Remembers sequence of events<br>- Summarizes story<br>- Predicts what will happen next<br>- Backs up statements with proof from reading<br>- Connects and builds to draw conclusions<br>- Uses prior knowledge to form an opinion<br>- Connects experiences with reading<br>- Evaluate/judges character, authors, books<br>• Verbally responds to literature in depth and is beginning to shift this ability to writing |
| **Skills/ Strategies** | • Identifies own name on print<br>• Understands "how" books work, e.g. top and bottom and front to back | • Recognizes that letters carry sounds<br>• Begins to use context, grammatical, and/or phonics cues and cross checks with pictures<br>• Matches words spoken to words in print<br>• Locates a known word<br>• Understands concepts about print, e.g. directionality, sentence, word, letter, space, beginning, end | • Increases and refines use of context, language, and/or phonics cues, and begins to use cross checking to self correct<br>• Begins to pause at appropriate places when reading orally<br>• Knows the meaning of a period, question mark, and exclamation mark<br>• Follows single step written instructions | • Uses a variety of ways to cross check and self correct<br>• Begins to read orally with expression and with appropriate pauses<br>• Knows the meaning of quotation marks and commas<br>• Follows two step written directions | • Self corrects automatically<br>• Confidently reads a story with appropriate expression<br>• Follows written multi-step directions<br>• Begins to ask questions about the structure of language |
| **Attitudes/ Behavior** | • Shows curiosity about print in environment<br>• Participates in the oral reading of familiar stories | • Is willing to read<br>• Focuses on print, supported by pictures<br>• Reading is vocal | • Selects books independently<br>• Shows familiarity with titles and authors<br>• Is beginning to read silently | • Chooses appropriate books to read for pleasure<br>• May choose books by author, topic, or a specific information<br>• Usually reads silently for an extended period of time, sometimes vocalizing when text is difficult<br>• Reads lengthier material | • Chooses to read a variety of materials for a variety of purposes<br>• Often chooses reading over other activities<br>• Reads silently for extended periods of time<br>• Recommends books to others |

Select the column(s) that best describe how a child habitually and naturally reads from a variety of materials at his/her *instructional reading level. (Instructional Level Material that is challenging but not frustrating with normal classroom instruction and support.)

*Juneau, Alaska School District Language Arts Portfolio Program, 1994*

BEST COPY AVAILABLE

# Project ALL Evaluation of Growth

**Funded through Title VII Bilingual Education**

**1991-92**

Student name ____________________

School ____________________

## CONTINUUM OF READING BEHAVIORS

1 2 3 4 5 **1** 1 2 3 4 5 6 7 8 9 **2** 1 2 3 4 5 6 7 8 9 **3** 1 2 3 4 5 6 7 8 9 **4** 1 2 3 4 5 6 7 8 9 **5** 1 2 3 4 5 6 7 8

### EMERGENT

- Unaware that print carries meaning
- Identifies own name in print
- Tells a story from pictures
- Answers questions about a teacher-read story
- Wants to hear favorite stories repeatedly
- Relies on memory for reading
- Joins in the oral reading of familiar stories especially those with rhythmic language, repetitive pattern
- Shows curiosity about print
- Uses comprehension strategies to respond to a teacher read story

### BEGINNING

- Focuses on print
- Recognizes that letters carry sounds
- Understands concepts about print, e.g. directionality, format of a book, sentence, word, line, letter, space, beginning, end
- Locates a given word in a known sentence
- Reads simple books in which language patterns are repeated
- Begins to use word identification strategies to identify words
- Can say another word that rhymes with a given word
- Begins to develop the meta cognitive awareness that reading must make sense and begins to apply cross checking strategies
- Begins to use comprehension strategies
- Shows some awareness to title, author, and illustrator
- Maintains independent reading for a short time, initially about 5 minutes

### DEVELOPING

- Reads known and predictable and other beginning books with growing confidence but still needs support with new and unfamiliar ones
- Develops further understanding of concepts of print (e.g. punctuation such as question marks, quotation marks )
- Increases and refines use of word identification strategies.
- Uses cross checking strategies to self correct
- Uses comprehension strategies more independently
- Chooses appropriate books to read for pleasure
- Maintains independent reading for a short time, initially about 10 minutes
- Shows familiarity with titles, authors, and illustrators

### INDEPENDENT

- Becomes an efficient silent reading (but sometimes vocalizes when text is difficult )
- After an initial reading, confidently reads a story to others with appropriate expression.
- Automatically applies multiple strategies to figure out words.
- Applies multiple strategies at the appropriate time to comprehend text
- Uses appropriate cross checking strategies
- Extends vocabulary with new words encountered in print.
- Selects challenging books independently.
- Attends to reading silently for an extended period of time initially about 15 minutes

### EXPANDING INDEPENDENT

- An efficient silent reader
- Reads aloud with expression and appropriate pauses.
- Adjusts strategies according to different reading material.
- Often reads Chapter books.
- Chooses to read a variety of materials for a variety of purposes.
- Identifies a variety of types of literature (i.e. biographies, nonfiction, poetry).
- Follows written multi-step directions.
- Summarizes a passage.
- Verbally responds to literature in depth and is beginning to in writing.
- is gaining understanding of elements of fiction, plot, characters, setting, point of view, and theme.
- Evaluates books making judgments, comparisons, analogies, and determines accuracy of information.
- Understands authors and illustrators have individual voices and styles.
- Recommends books to others.
- Often chooses reading over other activities.
- Reads silently for extended periods of time, often up to 30 minutes or more.

"Reading Strategies" on Reading Summary Sheet cross reference with these behaviors, habits, attitudes, and skills.

*Juneau, Alaska School District Language Arts Portfolio Program, 1994*

BEST COPY AVAILABLE

# JUNEAU PRIMARY LANGUAGE ARTS PORTFOLIO
1993-94

**Name:** _______________
**Teacher:** ______________

## CONTINUUM OF SPEAKING BEHAVIORS

| | Beginning | Developing | Expanding |
|---|---|---|---|
| Orally sharing information/experiences in small/large group | • Relates some of own experiences in a small group. | • Frequently shares own experiences in a small group. | • Orally shares information and own experiences in a small group. |
| | • Orally shares some of own experiences in a large group. | • Frequently shares own experiences in a large group. | • Consistently relates oral information and experiences in a large group. |
| Topic maintenance | • Often speaks of unrelated matters in a conversation. | • Usually maintains the topic in a conversation. | • Maintains the topic in a conversation and expands on the subject. |
| Vocabulary | • Uses a basic, core vocabulary that includes many non-specific words. | • Vocabulary is becoming more specific; beginning to make new words his/her own by generalizing to new situations. | • Uses a variety of words across many settings and contexts, hears words in class and readily makes them his/her own. |
| Retelling a story | • Retells a story relating some main events. | • Uses sequence and detail to retell a story and its main events. | • Uses inference and experiences with narratives to retell a story. |
| Asking questions | • Obtaining information by questioning is a struggle; still exploring structure of questions. | • Asks wh- questions; is developing precision in technique. | • Obtains necessary information by asking questions in a variety of situations with peers and adults. |
| Turn taking | • Learning when to talk and when not to talk in classroom and conversations. | • Frequently knows when to talk and when not to talk in conversation and classroom. | • Consistently knows when to talk and when not to talk in conversations and classroom. |
| Answering questions | • Learning to use question words and structuring. Some answers seem irrelevant. | • Answers wh- questions. | • Responds to variety of questions of increasing complexity. |
| Expressive voice | • Occasionally uses expression in voice. | • Words are intelligible; may be developing a fluent style. | • Uses a clear, fluent voice; varies it to situation. |
| Adaptation to listeners | • Occasionally shows awareness of audience while speaking. | • Developing a sense of audience. | • Adapts speaking to audience. |

BEST COPY AVAILABLE

Student name ______________________ Date ______________

# JUNEAU PRIMARY WRITING RUBRIC

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| IDEAS AND CONTENT | • **Communicates meaning through (single) picture(s).**<br>• Draws/scribbles as writing.<br>• Writes random letters or letter strings. | • **Communicates meaning with pictures supported by a word/ label.**<br>and/or<br>• Communicates a story through a related series of pictures.<br>• Writes names of people/things or environmental print. | • **Begins to communicate meaning with words.**<br>• Communicates meaning to reader through connected words. | • **Begins to write simple sentences, expressing a complete thought.**<br>• Ideas may be expressed in lists or patterned statements. | • **Begins to develop story sense.**<br>• Writer's emotion/ enthusiasm begins to show.<br>• Displays a sense of beginning and ending.<br>• Sequence of ideas is logical.<br>• Uses specific, accurate, suitable vocabulary. | • **Begins to add supportive details.**<br>• Generally stays on topic.<br>• Writer's emotion/ enthusiasm shows.<br>• Occasionally uses creative or vivid words/ phrases appropriately.<br>• Story has a beginning and an end. | • **Writes in a convincing, natural style.**<br>• Represents one main, focused idea/topic.<br>• Uses detail(s) to develop idea.<br>• Holds reader's attention.<br>• Organizes story with clear beginning/middle/ end.<br>• Communicates feelings to reader.<br>• Paints a picture in reader's mind.<br>• Experiments with uncommon words and new and/or different ways of saying things. |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| CONVENTIONS | • Writes shapes that look like letters. | • Uses mostly upper case letters. | • **Writes some recognizable, understandable words.**<br>• Attempts phonetic, temporary/invented spelling (using consonant beginning and/or ending sounds).<br>• Writes left to right/top to bottom.<br>• Begins to use spaces between words. | • **Writes simple sentences or pattern sentences.**<br>• May use repetitive sentence structure.<br>• Appropriately uses upper and lower case letters.<br>• Begins to use capitals and some punctuation correctly.<br>• Uses vowels in temporary/invented spelling.<br>• Begins to use conventional spelling.<br>• Consistently uses spaces between words. | • **Begins to use a variety of sentence structures.**<br>• Generally uses capitals and periods correctly.<br>• Frequently uses conventional spelling on high frequency words. | • **A wide variety of sentence structures is used.**<br>• Begins to punctuate correctly.<br>• Consistently uses conventional spelling on most words. | • **Uses varied and complex sentence syntax (construction of word, phrase, or clause in a sentence).**<br>• Writes clearly -- paper is easily understood.<br>• Consistently uses punctuation marks and capitalization correctly.<br>• Consistently uses conventional spelling. |

NOTE: You can use this rubric with some forms of writing, particularly fiction and non-fiction. If a paper contains only the child's name, it should not be factored into conventions. Pictures may be present at any level.

Revised 9/94

BEST COPY AVAILABLE

# VI-11

## Mathematics Problem Solving

OREGON DEPARTMENT OF EDUCATION

This set of performance criteria is used in the Oregon State Assessment to measure mathematical problem solving in grades 3, 5, 8, and 11. Students are given short, open-ended questions to solve individually. Students write down their answers and solution strategies. All papers are scored twice using the same generic scoring guides.

Contact: Wayne Neuberger
Assessment Supervisor
Office of Assessment
Public Services Building
255 Capitol St. NE
Salem, Oregon 97310-0203
Phone: (503) 378-8004

Copyright © 1994 by the Oregon Department of Education, Public Services Building, 255 Capitol St., N.E., Salem, Oregon 97310-0203. Phone: (503) 378-8004.

540

# MATHEMATICS PROBLEM SOLVING SCORING GUIDE

## Dimensions of Mathematics

### CONCEPTUAL UNDERSTANDING: Scoring Guide

Conceptual Understanding includes the ability to interpret the problem and select appropriate information to apply a strategy for solution. Evidence is communicated through making connections between the problem situation, relevant information, appropriate mathematical concepts and logical/reasonable responses.

5 **FULL CONCEPTUAL UNDERSTANDING:** The student uses all relevant information to solve the problem.
- The student's answer is consistent with the question/problem.
- The student is able to translate the problem into appropriate mathematical concepts.

3 **PARTIAL CONCEPTUAL UNDERSTANDING:** The student extracts the "essence" of the problem, but is unable to use this information to solve the problem.
- The student is only partially able to make connections between/among the concepts.
- The student's solution is not fully related to the question.
- The student understands one portion of the task, but not the complete task.

1 **LACK OF CONCEPTUAL UNDERSTANDING:** The student's solution is inconsistent or unrelated to the question.
- The student translates the problem into inappropriate mathematical concepts.
- The student uses incorrect procedures without understanding the concepts related to the task.

### PROCEDURAL KNOWLEDGE: Scoring Guide

Procedural Knowledge deals with the student's ability to demonstrate appropriate use of concepts. Evidence includes the verifying and justifying of a procedure using concrete models, or the modifying of procedures to deal with factors inherent in the problem.

5 **FULL USE OF APPROPRIATE PROCEDURES:** The student uses principles efficiently while justifying the solution.
- The student uses appropriate mathematical terms and strategies.
- The student solves and verifies the problem.
- The student uses mathematical principles and language precisely.

3 **PARTIAL USE OF APPROPRIATE PROCEDURES:** The student is not precise in using mathematical terms, principles, or procedures.
- The student is unable to carry out a procedure completely.
- The process the student uses to verify the solution is incorrect.

1 **LACKS USE OF APPROPRIATE PROCEDURES:** The student uses unsuitable methods or simple manipulation of data in his/her attempted solution.
- The student fails to eliminate unsuitable methods or solutions.
- The student misuses principles or translates the problem into inappropriate procedures.
- The student fails to verify the solution.

## PROBLEM SOLVING SKILLS AND STRATEGIES: Scoring Guide

Problem Solving requires the use of many skills, often in certain combinations, before the problem is solved. Students demonstrate problem solving strategies with clearly focused, good reasoning that leads to a successful resolution of the problem.

5 **EVIDENCE OF THOROUGH/INSIGHTFUL USE OF SKILLS/STRATEGIES: The skills and strategies show some evidence of insightful thinking to explore the problem.**
- **The student's work is clear and focused.**
- **The skills/strategies are appropriate and demonstrate some insightful thinking.**
- **The student gives possible extensions or generalizations to the solution or the problem.**

3 **EVIDENCE OF ROUTINE OR PARTIAL USE OF SKILLS/STRATEGIES: The skills and strategies have some focus but clarity is limited.**
- **The student applies a strategy which is only partially useful.**
- **The student 's strategy is not fully executed.**
- **The student starts the problem appropriately, but changes to an incorrect focus.**
- **The student recognizes the pattern or relationship, but expands it incorrectly.**

1 **LIMITED EVIDENCE OF SKILLS/STRATEGIES: The skills and strategies lack a central focus and the details are sketchy or not present.**
- **The procedures are not recorded (i.e., only the solution is present).**
- **Strategies are random.**
- **The student does not fully explore the problem looking for concepts, patterns or relationships.**
- **The student fails to see alternative solutions that the problem requires.**

## COMMUNICATION: Scoring Guide

In assessing the student's ability to communicate, particular attention should be paid to both the meanings he/she attaches to the concepts and procedures and also to his/her fluency in explaining, understanding, and evaluating the ideas expressed.

5 CLEAR, COMPLETE COMMUNICATION: The student gives a complete response with clear, coherent, unambiguous, and elegant explanations.
- The student communicates his/her thiking effectively to the audience.
- The details fit and make sense.
- One step flows to the next and shows organization.
- The student presents strong supporting arguments.

3 PARTIAL OR INCOMPLETE COMMUNICATION: The student's explanation is unclear, inconsistent or not complete.
- The student uses terminology incorrectly or inconsistently.
- The student's visual aids (graphs, tables, diagrams, etc.) are inappropriate or not directly related.
- The student's explanation centers on his/her solution, not on his/her thinking.

1 LIMITED OR LACK OF COMMUNICATION: The student's explanation is not understandable or not present.
- The student either does not use or misuses appropriate mathematical terminology.
- The student does not use essential visual aids to enhance or clarify the explanation.
- The student's explanation lacks focus.

## MATHEMATICS PROBLEM SOLVING SCORING GUIDE

| **Conceptual Understanding** - includes the ability to interpret the problem and select appropriate information to apply a strategy for solution. Evidence is communicated by making connections between the problem situation, relevant information, appropriate mathematical concepts, and logical/reasonable responses. | **Procedural Knowledge** - deals with the student's ability to demonstrate appropriate use of concepts. Evidence includes the verifying and justifying of a procedure using concrete models, or the modifying of procedures to deal with factors inherent in the problem. | **Problem Solving Skills** - require the use of many skills, often in certain combinations, before the problem is solved. Students demonstrate problem solving strategies with clearly focused, good reasoning that leads to a successful resolution of the problem. | **COMMUNICATION** In assessing the student's ability to communicate, particular attention should be paid to both the meanings s/he attaches to the concepts and procedures and also to his/her fluency in explaining, understanding, and evaluating the ideas expressed. |
|---|---|---|---|
| **5 Full Conceptual Understanding: The student uses all relevant information to solve the problem.**<br>• The student's answer is consistent with the question/problem.<br>• The student is able to translate the problem into appropriate mathematical concepts. | **5 Full Use of Appropriate Procedures: The student uses principles efficiently while justifying the solution.**<br>• The student uses appropriate mathematical terms and strategies.<br>• The student solves and verifies the problem.<br>• The student uses mathematical principles and language precisely. | **5 Evidence of thorough/insightful use of Skills/Strategies: The student's work demonstrates originality.**<br>• The student's work is clear and focused.<br>• The skills/strategies are appropriate and demonstrate some insightful thinking.<br>• The student gives possible extensions or generalizations to the solution or the problem. | **5 Clear Communication: The student gives a complete response with a coherent, unambiguous, and elegant explanation.**<br>• The student communicates his/her thinking effectively to the audience.<br>• The details fit and make sense.<br>• One step flows to the next and shows organization.<br>• The student presents strong supporting arguments. |
| **3 Partial Conceptual Understanding: The student extracts the "essence" of the problem, but is unable to use this information to solve the problem.**<br>• The student is only partially able to make connections between/among the concepts.<br>• The student's solution is not fully related to the question.<br>• The student understands one portion of the task, but not the complete task. | **3 Partial Use of Appropriate Procedures: The student is not precise in using mathematical terms, principles, or procedures.**<br>• The student is unable to carry out a procedure completely.<br>• The process the student uses to verify the solution is incorrect. | **3 Evidence of Routine or Partial Use of Skills/Strategies: The skills and strategies have some focus but clarity is limited.**<br>• The student applies a strategy which is only partially useful or partially executed.<br>• The student starts the problem appropriately, but changes to an incorrect focus.<br>• The student recognizes the pattern or relationship, but expands it incorrectly. | **3 Partial or Incomplete Communication: The student's explanation is unclear, inconsistent or not complete.**<br>• The student uses terminology incorrectly or inconsistently.<br>• The student's visual aids (graphs, tables, diagrams, etc.) are inappropriate or not directly related.<br>• The student's explanation centers on his/her solution, not on his/her thinking. |
| **1 Lack of Conceptual Understanding: The student's solution is inconsistent or unrelated to the question.**<br>• The student translates the problem into inappropriate mathematical concepts.<br>• The student uses incorrect procedures without understanding the concepts related to the task. | **1 Lacks Use of Appropriate Procedures: The student uses unsuitable methods or simple manipulation of data in his/her attempted solution.**<br>• The student fails to eliminate unsuitable methods or solutions.<br>• The student misuses principles, or translates the problem into inappropriate procedures.<br>• The student fails to verify the solution. | **1 Limited Evidence of Skills and Strategies: The student's work lacks a central focus or the details are sketchy or not present.**<br>• The procedures are not recorded (i.e., only the solution is present).<br>• Strategies are random.<br>• The student does not fully explore the problem looking for concepts, patterns or relationships.<br>• The student fails to see alternative solutions that the problem requires. | **1 Limited or Lack of Communication: The student's explanation is not understandable or not present.**<br>• The student either does not use or misuses appropriate mathematical terminology.<br>• The student does not use essential visual aids to enhance or clarify the explanation.<br>• The student's explanation lacks focus. |

BEST COPY AVAILABLE

# VI-12

# Group Performance Tasks: History-Social Science, Grade 11 Historical Investigations

California Learning Assessment System

In group performance assessment tasks, students combine knowledge of history with critical thinking, group participation, and oral and written communication. Classes work in small groups on various aspects of an issue or theme. Using this investigation, students offer responses or solutions using a historical perspective. The collaborative process can be assessed collectively and includes some individual writing.

The class is divided into small groups to analyze primary and secondary source materials that reflect different points of view. After reading and discussing the documents, each group decides on a position that it will share orally, either within their small group or with the entire class.

The task culminates with written responses to short-answer or essay prompts, which bring together prior knowledge about the general topic with what students learned during the task. Scoring criteria focus on success in group and collaborative learning, quality of thinking, effective communication of ideas, and knowledge and use of history. The rubric should be shared with the students.

Ideal tasks, designed for one or two days, should (1) be based on History-Social Science content, (2) require higher-order thinking skills, (3) involve discussion and activities by groups, and (4) include individually written exercises.

Contact: Samia Zumout
California Learning Assessment System
California Department of Education
P.O. Box 944272
Sacramento, California 94244
Phone: (916) 657-3011

---

Reprinted with permission from the California Department of Education, 721 Capitol Mall, P.O. Box 944272, Sacramento, California 97244-2720. Phone: (916) 657-3011

### California Assessment Program—History-Social Science Grade 11 Scoring Guide: Group Performance Task

| | Level I<br>Minimal Achievement | Level II<br>Rudimentary Achievement | Level III<br>Commendable Achievement | Level IV<br>Superior Achievement | Level V<br>Exceptional Achievement |
|---|---|---|---|---|---|
| Group and Collaborative Learning 20 | (1-4)<br>Exclusive reliance on one spokesperson<br>Little interaction<br>Very brief conversations<br>Some students are disinterested or distracted | (5-9)<br>Strong reliance on spokespersons<br>Only one or two persons actively participate<br>Sporadic interaction<br>Conversation not entirely centered on topic | (8-12)<br>Some ability to interact<br>At least half the students confer or present ideas<br>Attentive reading of documents and listening<br>Some evidence of discussion of alternatives | (13-16)<br>Students show adeptness in interacting<br>At least 3/4 students actively participate<br>Lively discussion centers on the task | (17-20)<br>Almost all students enthusiastically participate<br>Responsibility for task is shared<br>Students reflect awareness of others' views and opinions and include references to other opinions or alternatives in presentation and answers<br>Questions and answers illustrate forethought and preparation |
| Critical Thinking 30 | (1-6)<br>Demonstrates little understanding and only limited comprehension of scope of problem or issues<br>Employs only the most basic parts of information provided<br>Mixes fact and opinion in developing a viewpoint<br>States conclusion after hasty or cursory look at only one or two pieces of information<br>Does not consider consequences | (7-12)<br>Demonstrates only a very general understanding of scope of problem<br>Focuses on a single issue<br>Employs only the information provided<br>May include opinion as well as fact in developing a position<br>States conclusion after limited examination of evidence with little concern for consequences | (13-18)<br>Demonstrates a general understanding of scope of problem and more than one of the issues involved<br>Employs the main points of information from the documents and at least the general idea from personal knowledge to develop a position<br>Builds conclusion on examination of information and some consideration of consequences | (19-24)<br>Demonstrates clear understanding of scope of problem and at least two central issues<br>Uses the main points of information from the documents and personal knowledge that is relevant and consistent in developing a position<br>Builds conclusion on examination of the major evidence<br>Considers at least one alternative action and the possible consequences | (25-30)<br>Demonstrates a clear, accurate understanding of the scope of the problem and the ramifications of the issues involved<br>Employs all information from the documents and extensive personal knowledge that is factually relevant, accurate, and consistent in the development of a position<br>Bases conclusion on a thorough examination of the evidence, an exploration of reasonable alternatives and an evaluation of consequences |
| Communication of Ideas 20 | (1-4)<br>Position is vague<br>Presentation is brief and includes unrelated general statements<br>Overall view of the problem is not clear<br>Statements tend to wander or ramble | (5-9)<br>Presents general and indefinite position<br>Only minimal organization in presentation<br>Uses generalities to support position emphasizes only one issue considers only one aspect of problem | (8-12)<br>Takes a definite but general position<br>Presents a somewhat organized argument<br>Uses general terms with limited evidence that may not be totally accurate<br>Deals with a limited number of issues<br>Views problem within a somewhat limited range | (13-16)<br>Takes a clear position<br>Presents an organized argument with perhaps only minor errors in the supporting evidence<br>Deals with the major issues and shows some understanding of relationships<br>Gives consideration to examination of more than one idea or aspect of the problem | (17-20)<br>Takes a strong, well-defined position<br>Presents a well-organized, persuasive argument with accurate supporting evidence<br>Deals with all significant issues and demonstrates a depth of understanding of important relationships<br>Examines the problem from several positions |
| Knowledge and Use of History 30 | (1-6)<br>Reiterates one or two facts without complete accuracy<br>Deals only briefly and vaguely with concepts or the issues<br>Barely indicates any previous historical knowledge<br>Relies heavily on the information provided | (7-12)<br>Provides only basic facts with only some degree of accuracy<br>Refers to information to explain at least one issue or concept in general terms<br>Limited use of previous historical knowledge without complete accuracy<br>Major reliance on the information provided | (13-18)<br>Relates only major facts to the basic issues with a fair degree of accuracy<br>Analyzes information to explain at least one issue or concept with substantive support<br>Uses general ideas from previous historical knowledge with fair degree of accuracy | (19-24)<br>Offers accurate analysis of the documents<br>Provides facts to relate to the major issues involved<br>Uses previous general historical knowledge to examine issues involved | (25-30)<br>Offers accurate analysis of the information and issue<br>Provides a variety of facts to explore major and minor issues and concepts involved<br>Extensively uses previous historical knowledge to provide an in-depth understanding of the problem and to relate it to past and possible future situations |

BEST COPY AVAILABLE

# VI-13

## Thinking Skills Assessment

JEANNE BOND-ESSNER AND RACHEL KORACH
LAKE OSWEGO SCHOOL DISTRICT

The effort to find a better way to assess the thinking skills of high school students began in the 1990–91 school year. The current model has the following elements:

- A task which has intrinsic meaning for the students, causes them to engage in thinking, and stimulates them to produce a product which can be assessed. These will probably be multi-day projects and have portions done in a group.
- A written or presented performance product. These will be used both for course grades and for large-scale assessment.
- A generic, multi-trait scoring rubric.

For example, a task might be: "In the 'Adapt' project you learned about how a group of hunter/gatherers could use the resources in a given biome to develop a culture. What if the task was to create a farming/herding culture instead? Using the given environmental resources and characteristics, design a culture which uses farming/herding to meet people's needs."

Jeanne Bond-Essner and Rachel Korach
Principal Authors
Lake Oswego School District
P.O. Box 70
Lake Oswego, Oregon 97034-0070
Phone: (503) 636-7691.

---

Reprinted with permission from the Lake Oswego School District. Contributed by Ron Smith.

# DRAFT

## THINKING ASSESSMENT TRAITS
## LAKE OSWEGO SCHOOL DISTRICT

This model approaches the teaching of thinking as a combination of complex cognitive processes rather than as a series of discrete skills. The basic framework is founded on three fundamental concepts required for complex thinking processes:

- **differentiating**: the ability to delineate, to clarify, to define, to make distinctions, to use language precisely
- **distancing**: the ability to vary perspectives and to assume and to articulate multiple points of view while suspending judgment
- **designing**: the ability to construct meaning by developing and interrelating patterns of relationships through one's forming, structuring, ordering, and interpreting of experience

**Thinking as Differentiating**

5 The student reveals an accurate and complete knowledge and understanding of the topic. He/she reveals an understanding of the organization and structure of the material.

- Identifies, defines, and clarifies attributes of key concepts
- Makes clear distinctions between ideas or elements, appropriately separates topic into its component parts
- Establishes the relationship(s) of the elements to the whole; establishes a framework for essential concepts
- Communicates effectively through specific language, images, explanations, examples, etc.

3 The student reveals an incomplete or faulty knowledge and understanding of certain aspects of the topic. He/she has made an attempt to clarify and specify, but treatment is uneven or inadequate in some areas.

- Though some distinctions are accurately made, important differences are overlooked
- Though the topic is broken down into parts, important parts may be omitted, extraneous ones inserted, or inaccurate relationships presented
- There is inconsistent development of ideas
- Includes hidden assumptions or definitional problems
- Relies too much on generalizations or undefined terms
- An inadequate separation of the elements of the composition mar its organization
- Framework for presentation may be vague or inadequate

1 The student is ineffective in communicating knowledge or understanding of the topic or conveys inaccurate knowledge or misunderstandings of the topic.

- The presentation contains so many errors and inconsistencies that the whole becomes worthless
- Omits necessary identification or definitions of key elements
- Presents a hodge-podge of ideas; inappropriately lumps ideas or elements together without regard to important distinctions
- The topic is treated as a vague whole; no analysis of the topic is evident
- Instead of appropriately developing ideas, the student simply states--or restates--them
- Assumes too much, relies on vacuous generalizations, cliches, obvious "givens"

**Thinking as Distancing**

5 The student reveals an awareness of the purpose of the presentation, an awareness of the context, and an ability to view the problem/concept from multiple perspectives appropriate to that purpose and that context.

- Fulfills the purpose implied in the task and prompts
- Reveals an appropriate sense of audience
- Reveals an insightful understanding of a larger context
- Reveals an awareness and understanding of more than one viewpoint without losing the focus of the presentation; may consider alternative positions, show two or more sides of the issue, form final judgment from various perspectives, etc.

3 The student partially addresses the task but appears not to understand the purpose, the context, and/or complete intent of the prompt.

- Diction, tone, or choice of subject may reveal a poor sense of audience
- May state a knowledge of alternative viewpoints, but demonstrates a lack of understanding of them
- Too few sides of an issue are examined; focus may be inadequately narrow
- Reveals a limited awareness of context
- Strays from the subject somewhat

1 The student's response is inappropriate to the purpose and/or the context implied or stated in the task or prompt.

- Diction, tone, or choice of subject is not appropriate to the audience
- Reveals an inability to see or to understand multiple perspectives
- Reveals little or no awareness of context
- The presentation has no clear focus

**Thinking as Designing**

5 The student composes an original, coherent, inclusive, well-organized and meaningful new whole from the critical parts. He/she is in control of the topic.

- Reveals original thinking in the inferences, generalizations, interpretations, relationships, or synthesis presented
- Includes all the essential elements
- Structures clear, meaningful relationships among the concepts
- Engages the interest of the audience

3 The student combines the parts into a whole in a routine or incomplete manner.

- Though some original thinking is evident, the presentation relies heavily on structures, generalizations, interpretations, etc. which were presented in class
- The organization of the presentation may be unbalanced, ineffective, flawed
- The design makes sense but may be missing some key elements

1 The student has not combined the parts into an organized, coherent whole.

- The content of the presentation is a parroting back of material in the same form in which it was presented
- The organization of the presentation lacks a sense of beginning or ending; body lacks order or structure
- There is no discernible pattern, only a fragmented view of scattered information or ideas

BEST COPY AVAILABLE

# VI-14

# Assessing Portfolios

VICKI SPANDEL AND RUTH CULHAM
NORTHWEST REGIONAL EDUCATIONAL LABORATORY

This is a generic scoring guide that might be used with portfolios in any subject area. It might be most appropriately used with portfolios whose purpose is tracking student growth over time, accountability, student self-assessment, or diagnosis of student needs.

Vicki Spandel and Ruth Culham
Northwest Regional Educational Laboratory
101 S.W. Main, Suite 500
Portland, Oregon 97204
Phone: (503) 275-9500

---

Copyright © 1993 Vicki Spandel and Ruth Culham, NWREL.

# Change Over Time

## Strong

The student selects material that clearly demonstrates growth in one or more specific areas. Examples might be math problem solving skill or development of ideas in writing. In reviewing the examples, the reviewer can easily see how and in what areas the student has grown. The self-reflection (if present) generally matches what the reviewer sees.

## Developing

The samples show evidence of some growth, but the growth is limited. That is, the change from one sample to another is evident but not dramatic. Self-reflection, if present, may comment on changes that are not immediately apparent to the reviewer in looking at the student's work.

## Not Yet

The samples in the portfolio do not show evidence of noticeable student growth or change over time. Either noticeable growth has not occurred, or the student has not selected the samples of work that would illustrate that growth clearly.

Copyright Vicki Spandel and Ruth Culham, NWREL, 1993

# Diversity

## Strong

The portfolio clearly demonstrates that the student has tried a variety of tasks/projects/assignments/challenges. There is great variety in the kinds of work represented or the outcomes/skills demonstrated. For instance, a math portfolio might include some problem analysis, samples of graphing skill, a problem solving task that shows more than one solution, good use of math terminology, and a project showing application of math skills.

## Developing

The portfolio reflects some diversity. Tasks are not all parallel and do not all demonstrate identical outcomes. For instance, a math portfolio might include some open-ended problem solving with analysis of how the student did the task together with samples showing correct application of math procedures, concepts or symbols.

## Not Yet

The portfolio reflects minimal diversity. All tasks represented are more or less alike, and demonstrate the same outcomes/skills.

Copyright Vicki Spandel and Ruth Culham, NWREL, 1993

# Evidence of Problem Solving (Thinking/Reasoning)

## Strong

The work in the portfolio provides evidence that the student has identified, analyzed, planned strategies, and worked through the solution to a problem or question. For instance, a high school science student might tackle the question of how to preserve wetlands in a rapidly developing area. An elementary science student might consider the question of whether rats can thrive on a vegetarian diet.

## Developing

The work in the portfolio shows some evidence of thinking, reasoning, analyzing or problem solving, but the student may not have worked all the way through a solution, or may have missed opportunities to pull together interesting conclusions or plot alternate strategies. Still, the student's work overall shows signs of planning and purposeful effort.

## Not Yet

The student has not included in the portfolio any work that clearly demonstrates purposeful planning of strategies, problem solving, analysis of a situation, reasoning out a conclusion or considering alternative solutions.

Copyright Vicki Spandel and Ruth Culham, NWREL, 1993

BEST COPY AVAILABLE

# Organization, Format and Structure

## Strong

The student has formatted and arranged the portfolio in a way that invites the reader inside. Items within the portfolio are clearly labeled and dated; the sequence is purposeful. *All or most* of the following are included: a table of contents, a main title page or title pages for major sections, an introductory letter to reviewers, a statement of purpose (may be contained within the intro letter), criteria or rating scales (if relevant), a closing/summary comment or reflection.

## Developing

The portfolio is arranged and formatted in a way that enables the reader to make sense of it with a little work. Most items within the portfolio are labeled, dated or both. At least *some* of the following items are included: table of contents, letter of introduction, statement of purpose, title pages, criteria/rating scales, closing reflection.

## Not Yet

Arrangement and formatting of the portfolio make it difficult for the reviewer to determine when and under what circumstances it was assembled. Few items (if any) are clearly labeled or dated. Most or all of the following are missing: table of contents, letter of introduction, statement of purpose, title pages, criteria/rating scales, closing reflection.

Copyright Vicki Spandel and Ruth Culham, NWREL, 1993

BEST COPY AVAILABLE

# Self-Reflection

## Strong

Several (or more) examples of self-reflection show thoughtful consideration of personal strengths and needs based on indepth understanding of criteria. Reflections may also include a statement of personal goals; responses to learning, to a unit of study or to an assignment; a summary of growth over time; or other insight regarding the personal, individual story this student's portfolio tells.

## Developing

Self-reflections included within the portfolio provide at least a superficial analysis of strengths and needs, which may or may not be tied to specific criteria for judging performance or growth. The student may include comments on what he/she likes or dislikes about a content area or unit of study, or about what he/she finds difficult or challenging; but the reflections may not include insights regarding growth, needs, goals or changes in performance or learning styles over time.

## Not Yet

Either no self-reflection is included within the portfolio, or the self-reflection is rudimentary: e.g., "I put this in because I like it" "I included this in my portfolio because it took me a long time to do it" "This is in my portfolio because we were asked to put it in."

Copyright Vicki Spandel and Ruth Culham, NWREL, 1993

# VI-15

## Small-Group Discussion

ALBERTA EDUCATION, CANADA

The oral communication examination for English 30/33 (high school) includes a small-group discussion (3–4 students) of a story or other prose work or poem assigned as homework. Students are given several questions to discuss. The discussion takes one 60- to 80-minute period.

For example, one task on the 1991 assessment was to read "The Glass Roses" and discuss the following (among other things): "Decide what mood is developed by each description below and suggest reasons why these moods are effective in the context of the action in which they appear."

Students are assessed on Interaction, Comprehension Strategies, and Performance Skills.

Oral Communication Evaluation: English 30/33
activities and scoring guides
Alberta Education
Learning Resources Distributing Centre
12360 142 Street
Edmonton, Alberta T5L 4X9
Phone: (403) 427-2767

---

Reprinted with permission of Alberta Education, Devonian Building, West Tower, 11160 Jasper Avenue, Edmonton, Alberta, Canada TK5 OL2. Phone: (403) 427-0010.

# Group Skills

## SCORING GUIDE FOR SMALL GROUP DISCUSSION

(Talk will be characterized by pauses as student grapple with new ideas; the more thoughtful students may exhibit more pauses. False starts and repetitions will characterize talk until comprehension and linguistic control develop toward the end of the activity. Sentences may not be completed; students will interrupt one another, completing ideas begun by others.)

### INTERACTION

5 **Excellent:** The student is attentive, open-minded, courteous and sensitive to the ideas, tone and purpose of the activity; intellectual curiosity, attention to the task, and sensitivity to others help to create a productive climate in the group. The student confidently share ideas and feelings, actively builds on the ideas of others, and is an effective member or leader of the group.

4 **Proficient:** The students is attentive and courteous; purposefully and confidently undertakes the task assigned; thoughtfully handles the ideas offered by others, willingly offers own ideas, and is an efficient member or leader of the group.

3 **Satisfactory:** The student is courteous and willing to share with and listen to others; relies on the momentum of the group to motivate his inquiry into the topic; accepts ideas of others, offers own ideas, and is a courteous member or leader of the group.

2 **Limited:** The student is easily distracted; lacks the confidence to receive and express ideas easily and clearly; looks for confirmation of initial biases and may use language, tone, or nonverbal behavior inappropriate for the occasion or purpose; is a guarded, insecure member of the group and is uncomfortable in the role of leader.

1 **Poor:** The student is uninvolved in the activity and lacks the confidence to explore ideas orally; contributes almost nothing to discussions, which is inappropriate for the occasion and purpose; is uninvolved in the group and is unwilling to act as leader.

### COMPREHENSION STRATEGIES

5 **Excellent:** The student's discussion is perceptive, thorough, and insightful. Support is substantial and logical. The student shows perception while actively developing understanding of themes, main ideas, and supporting details; selects details from the text to support interpretations and revises interpretations to accommodate all details in the text; has a clear idea of the shape of the task and sustains inquiry until the task is thoroughly completed.

**4 Proficient:** The student's discussion is thoughtful and methodical. Support is appropriate and substantial. The student shows thoughtfulness as understanding of themes, main ideas, and supporting details develops; selects details from the text to support interpretations; has a clear idea of the shape of the task and sustains inquiry until the task is substantially completed.

**3 Satisfactory:** The students' discussion is on topic and sustained. Support is offered. The student shows a developing understanding of themes, main ideas, and supporting details; accepts ideas with little question but can explain and support ideas when asked by others to elaborate; has a mechanical understanding of the task and sustains inquiry until the task is completed.

**2 Limited:** The student's discussion is sporadic. The student struggles to provide ideas or support. The student has difficulty developing an understanding of themes and distinguishing between main ideas and supporting details; may frequently ask for repetition of ideas but shows little evidence of understanding; may have a limited understanding of the task or may understand the task but be unable to sustain the inquiry to adequately fulfill the task.

**1 Poor:** The student's discussion is weak and unfulfilling. The student does not develop an understanding of themes, main ideas, and supporting details; may passively accept ideas offered by others; involvement detracts from the activity rather than enhances it.

## PERFORMANCE SKILLS (VOCABULARY, LANGUAGE USE)

**5 Excellent:** The student uses precise vocabulary and economical syntax.
**4 Proficient:** The student uses precise vocabulary and expresses ideas clearly.
**3 Satisfactory:** The student uses general vocabulary and expresses ideas wordily.
**2 Limited:** The student uses vocabulary that is too general and immature to allow for adequate discussion beyond a fairly literal level.
**1 Poor:** The student's vocabulary and sentence skills are too deficient to support ideas required by the level of discussion.

---

Alberta Education, Alberta Canada, 1992.

# VII

## *Using Performance Assessment Information for Improvement*

# VII-1

# Using Performance Assessment Information for Improvement

## Introduction

Dean G. Arrasmith, James W. Kushman, and Nancey Olson
Northwest Regional Educational Laboratory

### Levels of Use for Performance Assessment Information

Performance assessment information is used for decision making at several levels. At the individual student level, specific and frequent decisions guide learning and instruction. Performance assessment information provides opportunities for teachers and students to understand the progress of learning and plan for subsequent instruction.

At the school and program levels, performance assessment information identifies effective teaching and learning opportunities, helps place students in appropriate instructional programs, and provides accountability information to constituent groups. For accountability, emphasis is on aggregated performance for a group of students rather than individual performance.

Performance assessment information is also useful at the school systems level, including school districts, and state and national systems. Certification of student competency— reflecting what students know and can do—is an important function of an education system and is informed by performance assessment data. Final demonstrations of proficiency are required for certification of competency, graduation or promotion. Additionally, policy and accountability issues can be addressed by information emerging from student performance assessments.

Papers in this section are organized around these levels of performance assessment information use. Section IV of this book addresses instructional implications of performance assessments and alignment with assessment. This section focuses on monitoring and directing student learning.

## Individual Student Use of Performance Assessment Information

**John Winbury and Christine Evans** describe the three-year implementation of portfolio assessments in San Diego, California's Poway School District. They found that teachers used portfolios to fill out report cards. As teachers relied less on traditional text-based assessments, they became "uncertain of their own judgments." The portfolios provided validation of the teachers' assessments and student grading. Teachers also reported using portfolios in parent and student conferences to accurately present student progress. Following the conferences, individual student learning goals were set.

Studying the impact of teaching a six-trait analytic writing model, **Judith A. Arter, Vicki Spandel, Ruth Culham, and Jim Pollard** found that students directly taught traits of writing wrote better essays related to those traits than students who were not directly taught them. When instruction and assessment are aligned, the information from the assessment can be directly related back to the quality of learning and need for further instruction.

By focusing students' attention specifically on the criteria for good writing, instructing and modeling it, and teaching students to make use of it, the overall quality of writing improves. In this sense, performance criteria provide a context and focus for instruction within which learning can be directed. Information from performance assessments related to well-defined criteria provide a valuable aid in directing instruction.

## School and Program Use of Performance Assessment Data

Discussing the use of school profiles for guiding school improvement efforts, **Robert E. Blum and Jocelyn A. Butler** describe 10 steps where schools use data to set and monitor school improvement goals. The process, focused on indicators of student achievement, behavior, and attitudes, encourages a schoolwide, cooperative effort for improvement.

A central element in planning and implementing school improvement efforts is developing a school profile. Indicators of student achievement including traditional and performance assessments, attitudes, and social behavior are suggested for content of the profile.

Using the results of the profile, schools are encouraged to develop goals for improvement and prescriptions to meet the goals. The profiling process, repeated periodically, provides a longitudinal view of the progress toward school improvement goals.

Vermont involves teachers from across the state in developing writing and math portfolio assessments. **Douglas E. Harris and Joyce Stone** from the Franklin Northwest Supervisory Union describe how their district uses statewide data and district-generated data for various purposes. They focus their attention on linking assessment to the change process and school improvement:

> The purposes for assessment grow out of four fundamental processes which should be an ongoing part of our schools: gathering information, focusing learning, making informed decisions, and revisiting the missions. At the heart of each one of the processes is the growing demand that schools be accountable to their communities.

Harris and Stone believe that criterion-based performance assessment serves several purposes of assessment. They provide three examples of how data were used to (1) improve a foreign language program; (2) provide an opportunity for an elementary school to engage in reflective practice based on student outcomes and dialogue among educators, students, and parents; and (3) focus on school improvement, allocating resources, staff assessments, and long-term improvements.

**Nancey Olson** describes a program developed at Crater High School in Central Point, Oregon. In response to a desire to make the vocational education program more responsive to the needs of a highly technological society, an applied technology program was developed, and a new facility designed and installed to replace an outdated metals program. Simulation data were created to show how this program relates to the new state criteria for the state's Certificate of Initial Mastery, defining new student outcomes for the 21st century.

**Claude Goldenberg and Bill Saunders** illustrate how performance assessments are used to promote equity excellence in schools serving disadvantaged and minority students. Their example from an urban, largely Latino elementary school in Southern California, shows how challenging performance assessments are used to assess student performance across different achievement levels. The premise of the article is that high standards and more challenging assessments do not, by themselves, put low-achieving groups at a disadvantage, nor do they automatically lead to high achievement for all students. Rather, a critical factor is how performance assessments are used to provide the larger context for improvement in the school.

The authors show how performance data can be disaggregated to determine if low-achieving students are improving along with high-achieving students as the school goes through an improvement effort. While setting high standards and developing challenging performance assessments are critical factors in the improvement process, they are not enough. A number of other contextual conditions must be present to achieve equity in excellence, including staff participation in setting goals and developing assessment tasks, a desire to improve, goals and assessments that continue to evolve as data are analyzed and interpreted, using assessments as opportunities for staff reflection and learning, and leadership pressure and support. Taken together, these contextual factors result in a school that emphasizes improving student achievement at all levels. Challenging performance assessments become the information and feedback loop on how well the staff is doing.

## Educational System Use of Performance Data

The article by **Kevin Castner, Lorraine Costella, and Steve Hess** and the paper by **Louise F. Waynant** illustrate how performance assessment is becoming an important part of state reforms in Maryland and how such assessments guide district and school improvement efforts. **Dale Armstrong's** paper describes the move to portfolio assessment in the Edmonton, Alberta schools.

Castner, Costella, and Hess discuss performance assessment as part of district restructuring efforts in Frederick County, Maryland. The article demonstrates how assessments, such as senior projects, can become aligned with the essential curriculum and used to assess learner outcomes emphasizing thinking, problem solving, and communication. Such performance assessments are used in concert with more traditional criterion-referenced tests in order to shift the balance from assessing only prerequisite skills to assessing higher-order skills that prepare students for careers and responsible citizenship.

The authors discuss how results of performance assessments are shared with the public, with results disaggregated by gender, ethnicity, and socioecomomic status. School improvement teams meet to review the data, assess progress, and plan further improvements. In Frederick County, the local performance assessment program is similar in intent and design to Maryland's School Performance Assessment Program, allowing a tight integration of state and local results. The authors describe how state and local results complement each other and are used in tandem to identify progress towards state and local performance standards. The article includes illustrations of how performance data are displayed and disaggregated.

Waynant, in another example from Maryland, discusses performance assessment in Prince George's County and how the challenging new state accountability standards are changing assessment practices from the state to the classroom level. She describes how new performance assessments and their results are used in the reform process.

Maryland's state assessment program includes extensive performance assessments at all grade levels. These are reported in terms of students meeting different proficiency levels and achieving state-defined "satisfactory" and "excellent" standards. Because performance assessment is such a strong part of the state's accountability system, Prince George's County has similarly built performance components into its more traditional "multiple choice" criterion-referenced tests in reading, writing, language arts, and mathematics. Teachers are also encouraged to develop and use classroom-level performance assessments for formative evaluation. At the school level, results are reported as individual student scores as well as percent of students scoring at defined score point levels. Results are disaggregated by race and gender.

Systemwide results from both state and local assessments have led Prince George's County to identify significant changes that the district needs to make in curriculum, instruction, and other areas. Waynant's article includes samples of how performance data are summarized and displayed in relation to identified standards, and how data are disaggregated by race.

The Edmonton Public Schools began a districtwide portfolio assessment project in 1989. Dale Armstrong reports on progress up to 1993 and provides rich data samples from Edmonton's project. Documenting student achievement and growth information is done annually and portfolio information is added to this database. Alignment with district goals is an important consideration at the school level. The districtwide assessment program supports both school-and district-level needs.

Student portfolios reflect intended outcomes of the program of studies. The assessment activities are age- and grade-level appropriate. All stakeholders find the portfolio data useful and understandable. Examples of district data show how the district aggregates the individual student portfolio data.

## Summary

Performance assessments and well-defined criteria for judging performance can provide a useful and focused information system for planning and evaluating student learning. At the classroom level, individual student information is useful to teachers and students for setting and monitoring the achievement of learning goals and growth. Teachers use performance assessments to help establish grades and provide information about learning as they move beyond textbook-driven curriculum and assessment.

Aggregation of performance outcomes for groups of students monitor program effectiveness and school improvement, and provide student achievement accountability to school constituents. Performance assessments, when closely aligned with instruction and with clear criteria of performance, provide the necessary information links between instruction and learning for program planning.

For educational systems, performance assessments of high quality can provide instructionally linked information for effective policy and decision making. System change can be monitored in terms of what students know and can do. Further, accountability information can be put into a teaching or learning context through the inclusion of illustrative performance at various levels of achievement. Understanding writing achievement scores is enhanced when levels of achievement are illuminated with examples of student writing. The same is true for other disciplines.

Finally, the information from performance assessments can be coordinated across levels of use and need. Performance assessments provide a foundation of what students know and can do, are aggregated and reported in useful ways for various purposes. Individual performance, perhaps disaggregated by essential traits of the performance, provide insight into quality of student learning. Aggregated across groups of students, the same performances can provide program and systems levels information about the effectiveness of educational programs and accountability information to school constituents. By coordinating information at different levels of need, a unified process of decision and policymaking can be encouraged linking learning and instruction to program development and policymaking.

# VII-2

## Poway Portfolio Project

JOHN WINBURY AND CHRISTINE SOBRAY EVANS,
POWAY UNIFIED SCHOOL DISTRICT, SAN DIEGO, CALIFORNIA

The Poway Portfolio Project was a response to the new California State Language Arts Framework (1989). The shift from skills-based instruction to an integrated, literature-based curriculum caused teachers to look for new ways to teach as well as assess students. Old notions of subskill testing for spelling, grammar, reading comprehension, etc. were inadequate in a classroom where the focus was more holistic. We began working with portfolios to get a broader view of students.

We designed our project to focus on three distinct but related audiences: the students, the teacher, and the grade-level team. For the students, the portfolio serves as a mirror. It reflects their progress. It allows them to assess strengths and weaknesses, and become a partner in their own assessment. For the teacher, the portfolio provides more information for student assessment than a meaningless string of numbers in a grade book. It documents movement and growth. Additionally, portfolios allow teachers to give parents a real view of student ability based on actual samples of work. This work also shows how students compare with others in their class. For the grade-level team, portfolios enable teachers to assess their language arts program and to determine their own inservice needs.

In the first year, we worked with reading specialists and principals at 11 elementary schools to pilot the portfolio project. At each school site we met with grade-level teams to brainstorm the kinds of writing done in the classrooms and then to select categories of inclusion for the first year's portfolios. Suggestions were given to teachers about portfolio management in the classroom:

- There should be both a writing folder and a portfolio for each student.
- Selections for the portfolio should be made by the student from his/her writing folder.
- All work should be dated and classified by type of writing.
- Periodically, students select work from the writing folder to be included in the portfolio. Selections are not static; work can be moved in and out of the portfolio.
- Each portfolio includes some form of self-evaluation.
- Teachers are encouraged to include reading artifacts, such as logs, fluency charts, book reviews, responses to literature, etc.

In the second year, we narrowed the focus of our project to the 4th grade teams at four elementary schools. We met with these teams to do preliminary readings of the portfolios. From the student papers, a narrative rubric was developed and anchor papers were selected to illustrate each score point. Perhaps the most important thing to emerge from this process was the idea of shared standards, i.e., teachers began to negotiate what constitutes high-quality work so that a grade-level team within a school would be in alignment for assessing student work.

In the third year, the development of a new report card based on rubrics rather than traditional letter grades supported the use of portfolios in language arts. The report card rubrics called for a "collection of work" to demonstrate student proficiency. As a result, elementary teachers throughout the district began to experiment with portfolios.

Teachers reported that they used the portfolios in several ways. Portfolios were critical in filling out report cards. In the past, they had relied on publishers' tests and others forms of traditional assessment to generate grades. This gradual process of "de-skilling" made assessments of student achievement, and that they had more confidence in the grades they gave. Portfolios were also used during parent conferences to give parents an accurate picture of student progress relative to the work of other students in the classroom. In student conferences, students reviewed their work from the previous grading period and set personal learning goals.

In the spring of 1991, we field-tested a reading of portfolios to compare them with the information gained from standardized proficiency testing. We wanted to see if we could eliminate the considerable expense of the testing process. We also wanted to eliminate the loss of instructional time that is taken up by test preparation and administration. We finally hoped to eliminate the anxiety felt by students and teachers in facing these proficiency tests.

### LISTENING RUBRIC
### Assessment Scale
### (Kindergarten)

**Secure Listener**

- Responds immediately to oral decisions
- Focuses on speaker
- Maintains appropriate attention span
- Listens to others
- Is interactive

**Developing Listener**

- Follows oral directions
- Usually attentive to speaker and discussions
- Listens to others without interrupting

**Beginning Listener**

- Has difficulty following directions
- Relies on repetition
- Often inattentive
- Short attention span
- Often interrupts

Instructions given to the teachers for this reading were to have each student select three pieces of writing from his or her portfolio and to include with this a list of books read during the year. Using the report card rubric, papers were read and given an overall score for writing. A separate rubric was developed to use in scoring the papers for spelling. Reading scores were determined by teacher observation and the book list.

Later this summer, proficiency scores will be compared with portfolio reading scores. We hope to find that we gain as much information from teachers in their classrooms as we do from standardized testing.

Interest in portfolios continues to grow in the district. Teachers in math, civics, and social studies have begun the process of implementing portfolios in their classrooms as well. Ultimately, we hope to implement a language arts portfolio program for grades K-12.

Following are the reading and language arts rubrics used with students in kindergarten through grade 5.

### LISTENING RUBRIC
### Assessment Scale
### (Grade 1–5)

**Strong Listener**

- Responds immediately to oral directions
- Focuses on speaker
- Maintains appropriate attention span
- Listens to others
- Is interactive

**Capable Listener**

- Follows oral directions
- Usually attentive to speaker and discussions
- Listens to others without interrupting

**Developing Listener**

- Has difficulty following directions
- Relies on repetition
- Often inattentive
- Short attention span
- Often interrupts

SPEAKING RUBRIC

### Assessment Scale
### (Kindergarten)

**Secure Speaker**
- Confident speaker
- Speaks loudly, clearly and with expression
- Expresses ideas with elaboration and support
- Consistently makes relevant contributions to class discussions

**Developing Speaker**
- Competent speaker
- Speaks loudly and clearly
- Expresses ideas in complete sentences
- Takes part in class discussions and stays on topic

**Beginning Speaker**
- May be a reluctant speaker
- Needs to work on speaking skills (volume, clarity, eye contact)
- Rarely contributes to class discussions in a meaningful way

SPEAKING RUBRIC

### Assessment Scale
### (Grades 1–5)

**Secure Speaker**
- Confident speaker
- Speaks loudly, clearly and with expression
- Expresses ideas with elaboration and support
- Consistently makes relevant contributions to class discussions

**Developing Speaker**
- Competent speaker
- Speaks loudly and clearly
- Expresses ideas in complete sentences
- Takes part in class discussions and stays on topic

**Beginning Speaker**
- May be a reluctant speaker
- Needs to work on speaking skills (volume, clarity, eye contact)
- Rarely contributes to class discussions in a meaningful way

READING RUBRIC

### Assessment Scale
### (Kindergarten)

**6 Experienced Reader**

Developing fluency as evidenced by reading phrases and sentences; self-corrects and takes risks as a reader; reads silently; uses multiple strategies to derive meaning from materials; reads with expression; reads new and familiar stories with accuracy; chooses to read for pleasure.

**5 Beginning Reader**

Reads word by word; has command of letters and sounds; understands the sound-symbol relationship; begins to use context, picture clues, phonics and knowledge of language to unlock new words and derive meaning from familiar books; reads signs, captions and titles; consistently uses one-to-one correspondence and self-corrects; begins to read familiar stories and predictable literature with accuracy.

**4 Developing Reader**

Makes the connection between oral language and written material; reads own writing; reads patterned text from memory; knows most letters and sounds; "pretend reads" and retells familiar literature; beginning to have one-to-one correspondence with words and a developing awareness to self-correct; reads some words in isolation; uses memory and pictures to gain meaning from materials.

**3 Pre-Reader**

Enjoys being read to; participates in shared reading; memorizes and repeats oral language patterns but does not yet connect to print; knows some letters and sounds; understands book-sense (left to right, front to back, right side up).

**2 Emergent Reader**

Enjoys begin read to; looks at pictures in books but does not yet make the connection to print; exhibits limited attention span during shared reading and may not participate; limited experience with letters and sounds; inexperienced book-sense (left to right, front to back, right side up).

**1 Inexperienced Reader**

(Teacher supplies appropriate descriptors)

BEST COPY AVAILABLE

### READING RUBRIC
## Assessment Scale
## (Grade 1)

6 **Experienced Reader**

A confident and independent reader who chooses from a wide range of new and familiar material; able to derive meaning from print; able to make predictions and draw inferences without teacher support; self-corrects and takes risks as a reader; seeks pleasure and information through reading.

5 **Strong Reader**

A capable reader who approaches familiar material with confidence, but still needs support with unfamiliar materials; beginning to make predictions and draw inferences from books and stories read independently; chooses to read independently; self-corrects for meaning.

4 **Capable Reader**

Shows success when beginning to explore new literature, but still needs to return to a familiar range of materials; beginning to read independently and apply a variety of reading strategies; begins to self-correct.

3 **Developing Reader**

Has increasing confidence in reading familiar and predictable material independently; has growing ability to use picture clues, memory, language patterns, and illustrations to gain meaning from the material; developing a one-to-one correspondence with words; applies phonics skills to reading material.

2 **Limited Reader**

Does not have ability to read print independently; relies on having the material read aloud by the teacher; may still be unaware that meaning comes from the material; knows some letters and sounds; participates in shared reading.

1 **Emergent Reader**

(Teacher supplies appropriate descriptors.)

### READING RUBRIC
## Assessment Scale
## (Grades 2–5)

6 **Exceptional Reader**

Reads and understands advanced materials; enjoys pursuing own interests independently; can summarize concisely; makes inferences and uses text to support ideas; uses wide range of strategies to deal with difficult text.

5 **Strong Reader**

Reads and understands books appropriate for grade level and above; reads during silent reading; selects appropriate books independently; can summarize; makes inferences; has strategies to deal with unfamiliar material.

4 **Capable Reader**

Reads and understands books appropriate for grade level; reads during silent reading; selects appropriate books independently; can re-tell a story; begins to make inferences; uses reading strategies for meaning.

3 **Developing Reader**

Reads and understands shorter books with simple narratives; does not always read during silent reading; relies on rereading familiar books; needs help independently reading and understanding grade-appropriate books.

2 **Limited Reader**

Reads and understands very easy and familiar material; has difficulty with unfamiliar material; rarely chooses to read for pleasure; needs a great deal of support in all areas.

1 **Emergent Reader**

(Teacher supplies appropriate descriptors.)

BEST COPY AVAILABLE

### WRITING RUBRIC
## Assessment Scale
## (Kindergarten)

**A collection of work will demonstrate proficiency**

**6 Exceptional Writer**

Enthusiastic, fluent, chooses to write independently; may write multiple sentences; word boundaries secure; some conventions of print and sentence sense; occasionally uses correct spelling.

**5 Experienced Writer**

Confident, developing fluency, sometimes writes independently; uses print from the environment in a meaningful way; begins to use the conventions of print; understands the use of spacing to define word boundaries; writes complete thoughts; primarily uses inventive spelling and begins to correctly spell high-frequency words.

**4 Beginning Writer**

Writes word-by-word; writes with guidance; beginning to use spacing to define word boundaries; may use beginning, middle and ending letters to represent words to demonstrate a more sophisticated understanding of sound/symbol relationships; may demonstrate a beginning awareness of conventions of print.

**3 Developing Writer**

Understands that oral language can be written; words are usually represented by beginning sounds; other letters may appear which demonstrate an understanding of sound/symbol relationships; writes word-by-word, but spacing to define word boundaries is not observed; can read own writing.

**2 Pre-Writer**

Drawing conveys message; may copy letters, words or sentences from the environment; beginning to use familiar letters and words as labels; writes familiar words from memory.

**1 Inexperienced Writer**

Communicates using pictures; uses scribble writing or letter-like marks; text conveys no meaning to reader; may dictate sentence to teacher.

### WRITING RUBRIC
## Assessment Scale
## (Grade 1)

**A collection of work will demonstrate proficiency**

**6 Exceptional Writer**

Enthusiastic, fluent, writes independently; has command of many conventions of print; demonstrates sentence sense; may use a variety of techniques to engage reader; writing shows originality and an organizational plan.

**5 Experienced Writer**

Competent and confident; fluent writer; writes independently; begins to use conventions of print in spelling, punctuation and grammar; word boundaries secure; text conveys the message; begins to use some techniques to engage reader; writing reflects logical sequence.

**4 Beginning Writer**

Developing fluency; writes competently with guidance; text beginning to reflect logical sequence; uses inventive spelling to convey the message; generally observes word boundaries; writing contains more apparent meaning; sentence structure may be simple and repetitive.

**3 Developing Writer**

Evidence of text; uses inventive spelling almost exclusively; beginning to observe word boundaries; writing begins to convey a comprehensible message but may be difficult to read; completes sentence patterns meaningfully; may copy words, phrases and sentences from class-generated ideas.

**2 Pre-Writer**

Communicates using pictures; some evidence of text but may not be readable; evidence of directionality; evidence of sound/symbol relationships; may copy from the environment without understanding that text has meaning.

**1 Emergent Writer**

(Teacher supplies directions)

BEST COPY AVAILABLE

## WRITING RUBRIC
## Assessment Scale
## (Grades 2–5)

**6 Exceptional Writer**

Enthusiastic and reflective writer who enjoys pursuing writing independently; uses a wide range of techniques to engage the reader; writing is fully-developed, appears effortless, shows originality, liveliness, excitement, humor or suspense.

**Collection of work demonstrates**

- Purposeful organizational plan
- Clear sentence sense and variety
- Control over vocabulary choice and arrangement
- Few errors in conventions and spelling
- Sustained quality throughout each piece

**5 Strong Writer**

Competent and confident writer who uses a wide range of techniques to engage the reader.

**Collection of work demonstrates**

- Strong organizational plan
- Clear sentence sense and variety
- Selection of vocabulary appropriate for the writing
- Few errors in conventions and spelling

**4 Capable Writer**

Comfortable with writing; developing skill and confidence; begins to use techniques to engage reader.

**Collection of work demonstrates**

- Organizational plan
- Varied sentence structure, may include fragments and/or run-ons
- Beginning use of precise vocabulary (more telling than showing writing)
- Occasional errors in conventions and spelling

**3 Developing Writer**

Developing fluency and writes with some degree of confidence; lacks techniques to engage reader.

**Collection of work demonstrates**

- Evidence of organizational plan but may lose focus
- Simple or repetitive sentences, may include fragments and/or run-ons
- Few descriptive words
- Errors in convention and spelling, but is readable

**2 Limited Writer**

May be a reluctant writer; has limited experience in writing

**Collection of work demonstrates**

- Inconsistent organizational plan
- Vague or confusing statements
- Simple sentence structure, many fragments and/or run-ons
- Limited vocabulary
- Frequent errors in conventions and spelling interfere with readability

**1 Emergent Writer**

(Teacher supplies appropriate descriptors)

BEST COPY AVAILABLE

# VII-3

# The Impact of Training Students to Be Self-Assessors of Writing

JUDITH A. ARTER, VICKI SPANDEL, RUTH CULHAM, AND JIM POLLARD
NORTHWEST REGIONAL EDUCATIONAL LABORATORY

Wouldn't it be great if we could integrate assessment and instruction so that our assessment tools are also our instructional materials? This study attempted to document the power of this approach in the teaching of writing. For the last 10 years, schools, districts, and states have been using a six-trait analytical scoring assessment model for tracking and analyzing the quality of student writing. The six-trait model uses professional judgment to rate pieces of writing along six dimensions:

- *Ideas.* The heart of the message, the content of the piece, the main theme, together with all the details that enrich and develop that theme.
- *Organization.* The internal structure of a piece of writing, the thread of central meaning, the pattern that holds everything together.
- *Voice.* The writer coming through the words, the sense that a real person is speaking to us and cares about the message.
- *Word Choice.* The use of rich, colorful, precise language that communicates not just in a functional way but in a way that moves and enlightens the reader.
- *Sentence Fluency.* The rhythm and flow of the language, the sound of word patterns, the way the writing plays to the ear—not just to the eye.
- *Conventions.* The mechanical correctness of the piece—grammar, spelling, paragraphing (using indents), capitalization, and punctuation.

(The Appendix to this chapter shows a student version of the scoring guide. The criteria examples in Section VI show the scoring guide that adult raters use.)

In using the six-trait model for large-scale assessment, we have noticed that those doing the scoring seem to learn a lot about writing. In addition to learning about how students write, they become better writers themselves. They also report that they are better able to teach writing, because now they understand what they need to teach.

For the last five years, we have been training teachers to teach students to be assessors of writing (both their own and that of others) using the six traits. The rationale is that if raters of writing learn

---

Paper presented at annual meeting of the American Educational Research Association, New Orleans, 1994.

so much about writing from systematically analyzing and rating it, why not pass this power on to students? Anecdotal evidence from teachers strongly supports the power of this notion; teachers tell us how much more comfortable they feel teaching writing and how much better student writing has become since systematically gearing instruction to the six-trait model.

We conducted a study from November 1992 to April 1993 to collect evidence to support or refute the power of this approach. The specific research question we wanted to answer was "Does the writing of students who have direct instruction on assessing writing using the six-trait analytical model improve more than the writing of students who do not have such instruction?"

### Method*

To represent a range of learning contexts (e.g., rural and urban, size of district, size of school, and student expenditures) and student types, we recruited six classrooms of 5th graders to participate in the study. Classrooms were in different buildings and school districts and were randomly assigned to either a "treatment" (integration) or a "control" (observation) condition. Sixty-seven students (three classrooms) participated in the integration group and 65 students (three classrooms) in the observation group.

Teachers at the integration sites (plus any other teachers in the school who wished to come) attended a one-day training session in October 1992. The session focused on writing assessment and integrating the six-trait model into instruction. This training included definitions of the six traits, opportunity to practice assessing student writing for each trait, and instruction on how to teach the traits.

Teachers received an extensive set of materials to serve as a resource throughout the school year. Materials included scoring guides for both teachers and students at the 5th-grade level, plus sets of classroom activities to help students develop skills in each trait. In addition, teachers were given sample student essays to share and discuss with their students and strategies for successfully teaching students to become self-assessors using the six-trait model.

Northwest Regional Educational Laboratory staff visited each "integration" teacher eight times between November 1992 and April 1993 to help implement the strategies presented during the initial training session. For the first visit, staff led students through activities designed to provide an overview of the six traits and began training on the Ideas trait. The second and third visits continued training on Ideas. The fourth and fifth visits dealt with Organization and Voice, respectively. The final visit was a wrap-up and debrief. Staff mentioned the last three traits (Word Choice, Sentence Fluency, and Conventions) but did not address them in any depth.

At the integration sites, teachers provided students with extensive direct instruction on the traits. This instruction included (1) sharing and evaluating (as a group) samples of literature that exemplified each trait; (2) evaluating the work of students from other classrooms and discussing how to improve the writing on the trait under consideration; (3) using minilessons to teach skills for each trait (e.g., how to generate ideas, different kinds of organization, and how to write good leads); and (4) applying the traits to students' own writing (either self-assessment or peer review). Students received their own scoring guides. In one classroom, parents also received instruction in the model so that they could provide reinforcement at home.

Staff also visited each control teacher three times to determine how writing instruction occurred. Staff simply noted the kinds and amount of writing instruction that teachers provided and asked questions about normal strategies for writing instruction. All control sites used a process approach to writing: Students had an opportunity to write more than one draft and sometimes to share their writing with the teacher or with peers to obtain feedback. The amount of writing students did in the classroom in a week varied from 30 minutes to more than three hours. Staff observed that control teachers did not directly instruct students on the qualities of good writing; they did, however, provide instruction in prewriting, such as brainstorming and listing. In one classroom, students had opportunities to use word processing for structured assignments and to help each other with revising and editing. Students were sometimes instructed to do additional writing outside of class.

Thus, the writing instruction provided in the control and integration classrooms was substantially different. (Note: At the end of the study, teachers and other interested staff in the control schools also received a one-day training on writing assessment, and the packet of instructional materials.)

### Data Source and Study Design

Before and after the study, students wrote essays under the following conditions: They had three 45-minute periods to write an essay on an assigned topic; teachers allowed time to prewrite, create and read over a rough draft, revise, edit, and produce a final copy; students were permitted to use dictionaries or other written references during testing but could not confer with the teacher or with each other.

---

*Note to the nontechnical reader:* Please feel free to skip from here to the "Discussion" section. The next few pages deal with technical design and are included for those who care about such things.

Teachers used two prompts: expository to elicit informational writing; and narrative to elicit stories:

> **Expository:** Think of the most valuable thing you own which was not bought in a store. Explain why it is important to you.
>
> **Narrative:** You won't believe it, but here's what happened.

At pre-test time, teachers randomly distributed the two prompts in each classroom. At post-test time, students responded to the prompt they did not get at pre-test time. This procedure (1) eliminated memory effects that come from repeating prompts and (2) controlled for differences in difficulty of prompts. Random distribution reduced self-selection effects and ensured that, on average, student ability did not interact with the prompt.

We used pre-test information to plan instruction in the three treatment classrooms. Since these pre-tests were also used at the end of the study to measure impact on students, other raters rescored them along with the post-tests. For the impact study, we controlled for possible rater bias by removing paper identification, mixing pre- and post-test essays, and randomly distributing them to experienced adult raters. Each paper received scores on each trait from two raters.

Data were analyzed using six (one for each trait) repeated measures (matched pre- and post-test scores) analyses of variance.

### Results

Figure 1 shows pre-test and post-test means. The pre-test scores for the experimental (treatment) and control groups are very similar on all traits. The treatment group gained the most on those traits receiving the most emphasis in instruction. Performance on other traits gained slightly. Students in the control group gained slightly on two traits and remained essentially the same on the other four.

Figure 2 shows the statistical results. Gains between the experimental and control groups were significantly different for Ideas, approach significance for Organization, tended toward significance for Voice, and were nonsignificant for Word Choice, Sentence Fluency, and Conventions.

### Discussion

Student scores improved on traits in proportion to the amount of time spent on them and the order in which they were introduced. Thus, students showed the most improvement on Ideas, the trait introduced first; an almost significantly different gain on Organization, the second trait introduced; and a slightly lower gain on Voice, the third trait introduced. No significant difference occurred in gains on the three traits that were not directly taught to students (Word Choice, Sentence Fluency, and Conventions). This finding lends credibility to the premise that student writing improves to the extent that we address instruction to the features of writing we deem most important, and to the extent that we directly teach students what good writing and poor writing look like on each of these dimensions.

FIGURE 1

PRE-TESTS AND POST-TESTS MEANS FOR EXPERIMENTAL (TREATMENT) AND CONTROL GROUPS

| Trait | Group | Pre-Test | Post-Test |
|---|---|---|---|
| Ideas | Experimental | 2.54 | 3.38 |
| | Control | 2.68 | 2.75 |
| Organization | Experimental | 2.60 | 3.15 |
| | Control | 2.61 | 2.70 |
| Voice | Experimental | 2.73 | 3.60 |
| | Control | 2.91 | 3.12 |
| Word Choice | Experimental | 2.73 | 3.26 |
| | Control | 2.91 | 3.11 |
| Sentence Fluency | Experimental | 2.85 | 3.12 |
| | Control | 2.89 | 2.89 |
| Conventions | Experimental | 2.79 | 2.98 |
| | Control | 2.99 | 2.99 |

But we can hear some people objecting—"Of course students in the integration group do better than those in the control group; they were directly instructed on the traits used in the test. We know that direct instruction on the traits will improve scores. The real question is, Does writing improve?" The response to this poses the counter-question, "And how do we get this external measure of the extent to which writing improves?" The only possible way is to make a direct professional judgment of whether writing has improved. We contend that the six-trait model is the best available means of making this judgment. If the six-trait analytical model does define what we mean by good writing, then improvements on the traits is a *direct* measure of writing improvement. The direct, professional judgment of writing quality is what the six-trait model is all about. *No better outside measure exists.*

Therefore, this study does not involve "teaching to the test"; it involves "teaching to the criteria." Therein lies the power and the whole point of the technique of teaching clearly defined performance

BEST COPY AVAILABLE

FIGURE 2

ANALYSIS OF VARIANCE RESULTS

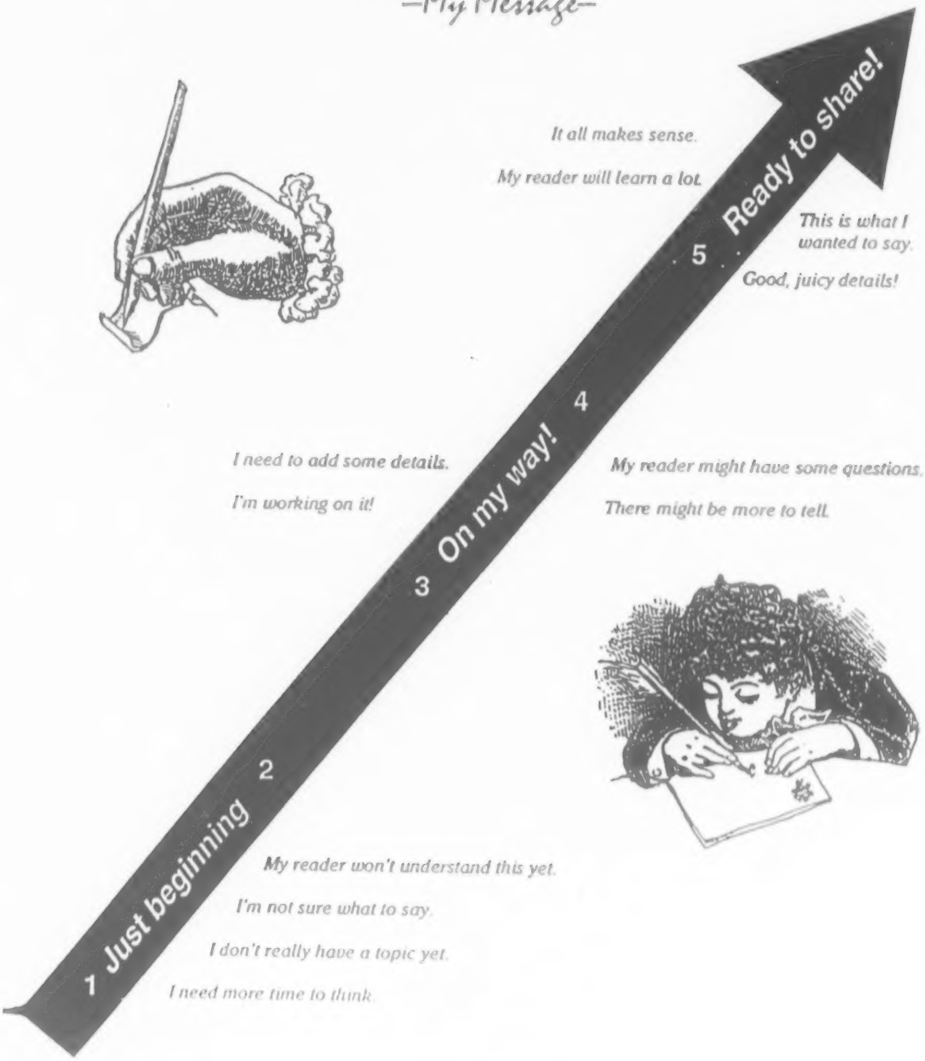
| Trait | Source of Variance | SS | DF | MS | F | Significance |
|---|---|---|---|---|---|---|
| Ideas | Within+Residual | 126.44 | 131 | .97 | 3.98 | .048 |
| | Group | 3.84 | 1 | 3.84 | | |
| Organization | Within+Residual | 128.23 | 131 | .98 | 3.08 | .082 |
| | Group | 3.01 | 1 | 3.01 | | |
| Voice | Within+Residual | 114.94 | 131 | .88 | 1.59 | .209 |
| | Group | 1.4 | 1 | 1.4 | | |
| Word Choice | Within+Residual | 60.85 | 131 | .46 | .04 | .839 |
| | Group | .02 | 1 | .02 | | |
| Sentence Fluency | Within+Residual | 117.29 | 131 | .90 | .67 | .414 |
| | Group | .60 | 1 | .60 | | |
| Conventions | Within+Residual | 177.79 | 131 | 1.36 | .53 | .466 |
| | Group | .72 | 1 | .72 | | |

criteria directly to students. If the six-trait model defines what we value in writing, then teaching students the traits teaches them what good writing is. This study tends to support the conclusion that student writing will improve if students receive this type of instruction. Therefore, we can affect student achievement by improving classroom assessment techniques, teacher skills in using them, and student self-assessment

## APPENDIX A
## SCORING GUIDE USED BY FIFTH GRADE STUDENTS

### IDEAS

—My Message—



*It all makes sense.*

*My reader will learn a lot.*

*This is what I wanted to say.*

*Good, juicy details!*

*I need to add some details.*

*I'm working on it!*

*My reader might have some questions.*

*There might be more to tell.*

*My reader won't understand this yet.*

*I'm not sure what to say.*

*I don't really have a topic yet.*

*I need more time to think.*

© Vicki Spandel and Ruth Culham, Center for Classroom Assessment, NWREL 1992

BEST COPY AVAILABLE

Appendix A—*(continued)*

## SCORING GUIDE USED BY FIFTH GRADE STUDENTS

# ORGANIZATION

*—From Beginning to End—*

*I know where I'm going.*

*I love my beginning.*

*My ending is cool.*

**5 Ready to share!**

*Follow me!*

*I see just how all the parts fit together.*

**4**

*I'm starting to see a pattern here!*

*My beginning is OK.*

**3 On my way!**

*My paper is PRETTY easy for a reader to follow.*

*Maybe I need to move some things around.*

*I'm not sure how to end it.*

*How do I begin?*

**2**

*What should I tell first?*
*What comes next?*

*My reader would feel LOST!*

**1 Just beginning**

*Help!*
*Which pieces go together?*

*I don't know where I'm headed.*

© Vicki Spandel and Ruth Culham, Center for Classroom Assessment, NWREL, 1992

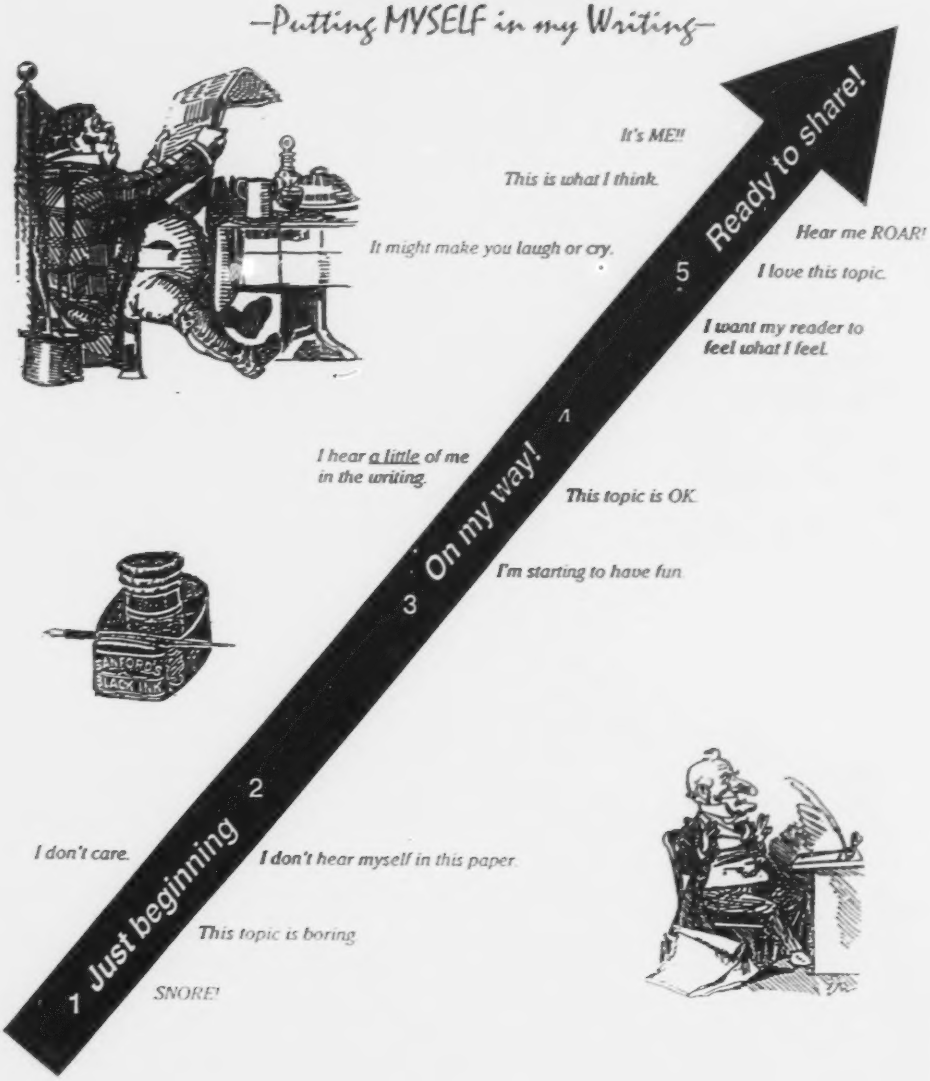
BEST COPY AVAILABLE

APPENDIX A—*(continued)*

## SCORING GUIDE USED BY FIFTH GRADE STUDENTS

# VOICE

*–Putting MYSELF in my Writing–*


It's ME!!
This is what I think.
It might make you laugh or cry.
Ready to share!
5
Hear me ROAR!
I love this topic.
I want my reader to feel what I feel.
4
I hear a little of me in the writing.
On my way!
This topic is OK.
I'm starting to have fun.
3
2
I don't care.
Just beginning
I don't hear myself in this paper.
This topic is boring.
1
SNORE!
SANFORD'S BLACK INK

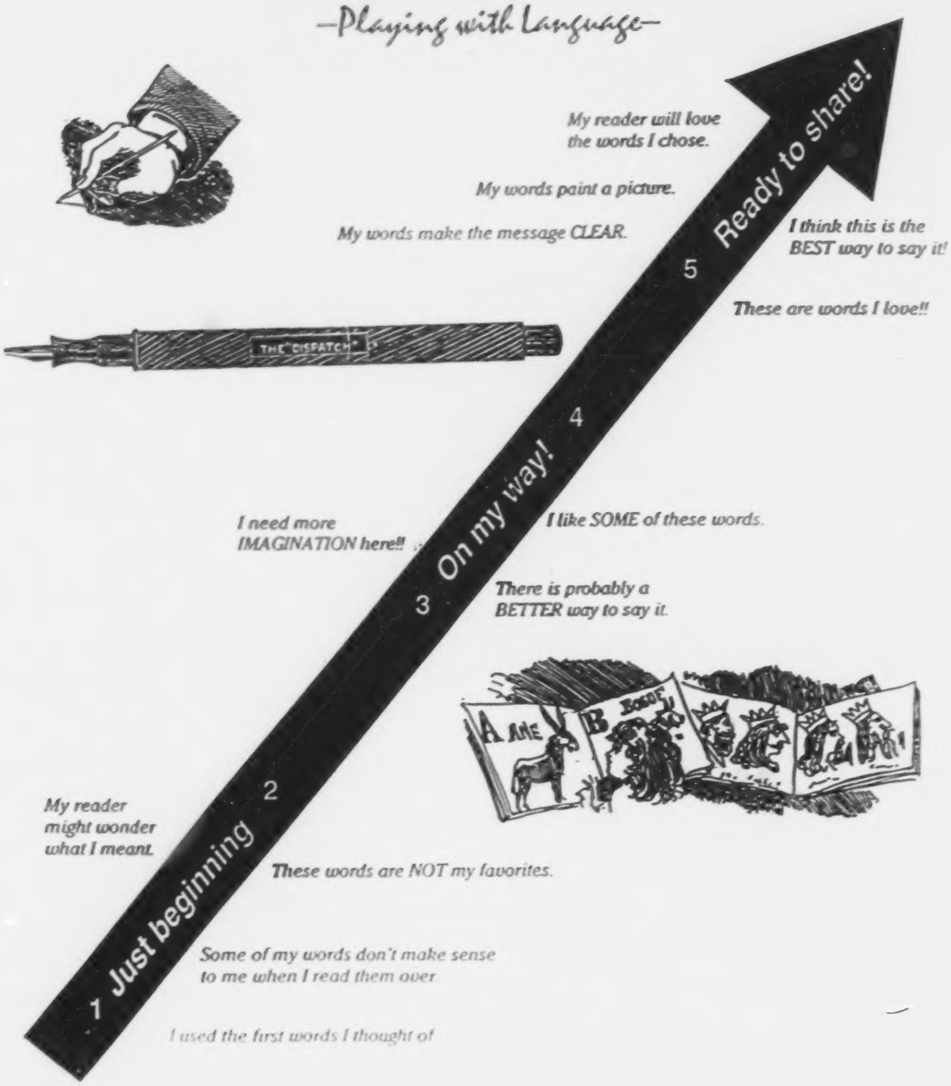
© Vicki Spandel and Ruth Culham, Center for Classroom Assessment, NWREL, 1992

APPENDIX A—*(continued)*

## SCORING GUIDE USED BY FIFTH GRADE STUDENTS

# WORDS

*—Playing with Language—*



*My reader will love the words I chose.*

*My words paint a picture.*

*My words make the message CLEAR.*

*I think this is the BEST way to say it!*

*These are words I love!!*

*I need more IMAGINATION here!!*

*I like SOME of these words.*

*There is probably a BETTER way to say it.*

*My reader might wonder what I meant.*

*These words are NOT my favorites.*

*Some of my words don't make sense to me when I read them over.*

*I used the first words I thought of.*

© Vicki Spandel and Ruth Culham, Center for Classroom Assessment, NWREL, 1992

APPENDIX A—*(continued)*

## SCORING GUIDE USED BY FIFTH GRADE STUDENTS

# FLUENCY

*—Listening to the Sound—*

5 Ready to share!

*My sentences make sense.*

*My paper is EASY to read out loud.*

*Some sentences are LONG and some are SHORT.*

*My sentences begin in several different ways.*

*I love the sound of this paper.*

4

3 On my way!

*A lot of my sentences begin the same way.*

*I wish my paper sounded a little smoother.*

*My paper is PRETTY easy to read out loud if you take your time.*

2

1 Just beginning

*Some of my sentences do not make sense.*

*My paper is HARD to read out loud—even for me!*

*Sometimes, I can't tell where to begin a new sentence.*

*I'm not happy with the sound of this paper yet.*

© Vicki Spandel and Ruth Culham, Center for Classroom Assessment, NWREL, 1992

APPENDIX A—*(continued)*

## SCORING GUIDE USED BY FIFTH GRADE STUDENTS

# CONVENTIONS

*—Editing—*

5 Ready to share!

*You have to look hard to find mistakes in my paper!*

*I used capitals to BEGIN SENTENCES.*

*My work has been edited and proofread; it is EASY to read.*

*I used capitals on NAMES.*

*I checked my spelling.*

*I used periods (.) and question marks (?) in the right spots.*

4

3 On my way!

*I'm not sure I spelled all the words right.*

*I might have left out a period or question mark.*

*My paper has some mistakes, but you can still read it.*

*I have more editing to do.*

2

1 Just beginning

*My reader would notice lots of mistakes.*

*I have not edited my work yet.*

*I need to correct some things before I publish this work.*

© Vicki Spandel and Ruth Culham, Center for Classroom Assessment, NWREL, 1992

# VII-4

## Managing Improvement by Profiling

ROBERT E. BLUM AND JOCELYN A. BUTLER

School improvements should be aimed directly at improvement of student performance—academic achievement, social behavior, and attitude—or they may miss the mark. The Goal Based Education Program at the Northwest Regional Educational Laboratory (NWREL) has developed a process called *profiling* that is used to gather information about student performance. The resulting profile can then be used as the basis for planning and implementing improvements.

Profiling is part of an overall school improvement program called Onward to Excellence. Schools in the Onward to Excellence program contract for training and technical assistance for a year to learn and implement an improvement process based on the results of effective schools research as synthesized by NWREL staff. The program provides a ten-step process whereby schools collegially use data to set goals for improvement:

1. Introduce the program to all school staff members, form a leadership team (the school principal, staff teachers, and a central office representative) to head efforts, and set the stage for improvements.
2. Study the effective schools research base and findings.
3. Profile current levels of student performance (academic achievement, social behavior, and attitude).
4. Set a goal for improvement.
5. Check current instructional practices.
6. Develop a research-based prescription to meet the improvement goal.
7. Prepare for implementation.
8. Implement the prescription.
9. Monitor implementation.
10. Evaluate progress and renew efforts.

All improvement efforts in the Onward to Excellence program focus on student performance: the indicators of school improvement are changes in student achievement levels, behavior, and attitude. NWREL's experience has shown that team-based improvements are effective in achieving the cooperative, schoolwide efforts necessary to plan and implement lasting change. Local schools develop their own resources for improvement: and the process becomes an ongoing approach for managing local school improvement efforts.

### Profiling a School

Profiling takes place in steps three and four of the process. In step three, school teams plan for profiling,

---

Adapted from R.E. Blum and J.A. Butler, "Managing Improvement by Profiling," *Educational Leadership* 42, 6 (March 1985): 54-58.

introduce the process to staff, collect and summarize schoolwide data, and prepare a written report, or *profile*, of the school. In step four, the school reviews the profile to identify strengths and weaknesses in student performance; set priorities for improvements; establish goals; and describe profile results and goals to the school community. Planning and implementation of improvements are then based on profiling results. (See Figure 1 for an overview of the profiling process.)

Profile preparation involves the two major tasks of collecting and summarizing information and preparing a written version of the findings. Both tasks involve a number of steps that should be managed through planning and assigning specific tasks to team members. In addition, it is crucial that all staff members be informed of the profiling effort and that information collection be focused on creating a picture of the school as a whole. Schoolwide data are needed, not to monitor individual teacher or student performance, but to bring about schoolwide improvement.

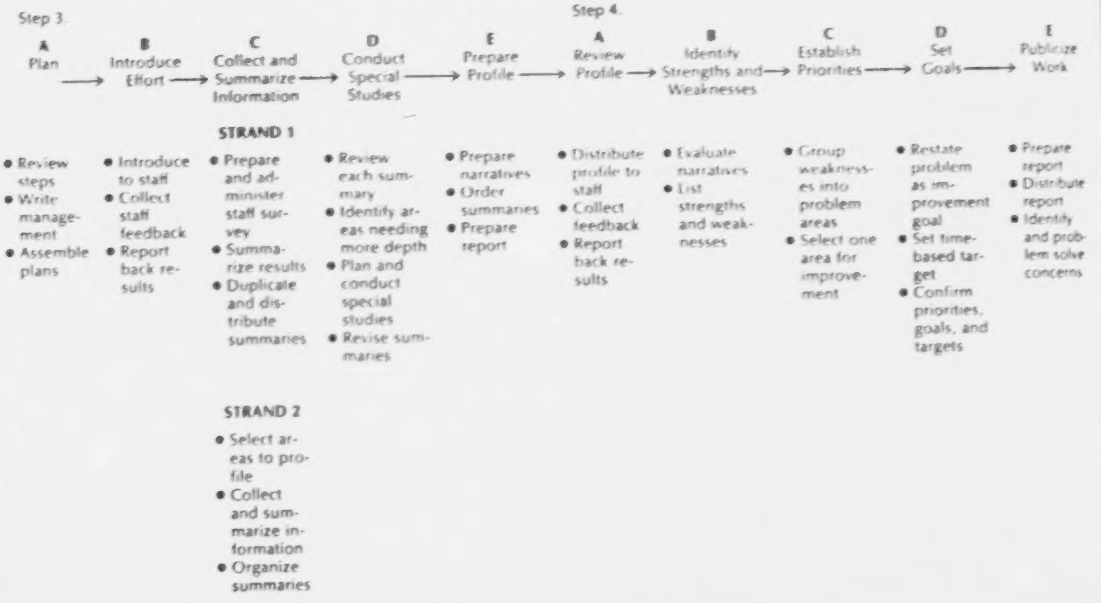
## Collecting and Summarizing Information

After the leadership team is formed and has become familiar with the overall thrust of the improvement program, they begin planning to collect current data on the school. Using a Profile Contents Menu (Figure 2), team members discuss types of information they will collect. The menu has three sections, each dealing with one area of student performance: academic achievement, student attitude, and social behavior. To achieve an overall view of students in the school, teams should collect at least one group of data in each of the three sections.

Using the Menu as a worksheet, teams check off data sources already available to them. Results of achievement tests, attendance records, and student attitude surveys, for example, are readily available in many schools. Team members then note areas where additional data would be useful. Using group decision-making techniques, the team considers available and desirable data and creates a firm list of items to collect for the profile. Team members are then assigned specific tasks for data collection, creating a plan for managing the profiling effort.

With the management plan and selected data sources in hand, the team presents the improvement program, the profiling process, and the profiling areas to the entire staff. This action both introduces the effort to the staff and provides a means to collect staff feedback on the profiling process. Following the management plan, the team collects and summarizes data in the selected profiling areas.

**Figure 1. Profiling Overview.**


Step 3.
A
Plan
B
Introduce Effort
C
Collect and Summarize Information
D
Conduct Special Studies
E
Prepare Profile
Step 4.
A
Review Profile
B
Identify Strengths and Weaknesses
C
Establish Priorities
D
Set Goals
E
Publicize Work
STRAND 1
• Review steps
• Write management
• Assemble plans
• Introduce to staff
• Collect staff feedback
• Report back results
• Prepare and administer staff survey
• Summarize results
• Duplicate and distribute summaries
• Review each summary
• Identify areas needing more depth
• Plan and conduct special studies
• Revise summaries
• Prepare narratives
• Order summaries
• Prepare report
• Distribute profile to staff
• Collect feedback
• Report back results
• Evaluate narratives
• List strengths and weaknesses
• Group weaknesses into problem areas
• Select one area for improvement
• Restate problem as improvement goal
• Set time-based target
• Confirm priorities, goals, and targets
• Prepare report
• Distribute report
• Identify and problem solve concerns
STRAND 2
• Select areas to profile
• Collect and summarize information
• Organize summaries

BEST COPY AVAILABLE

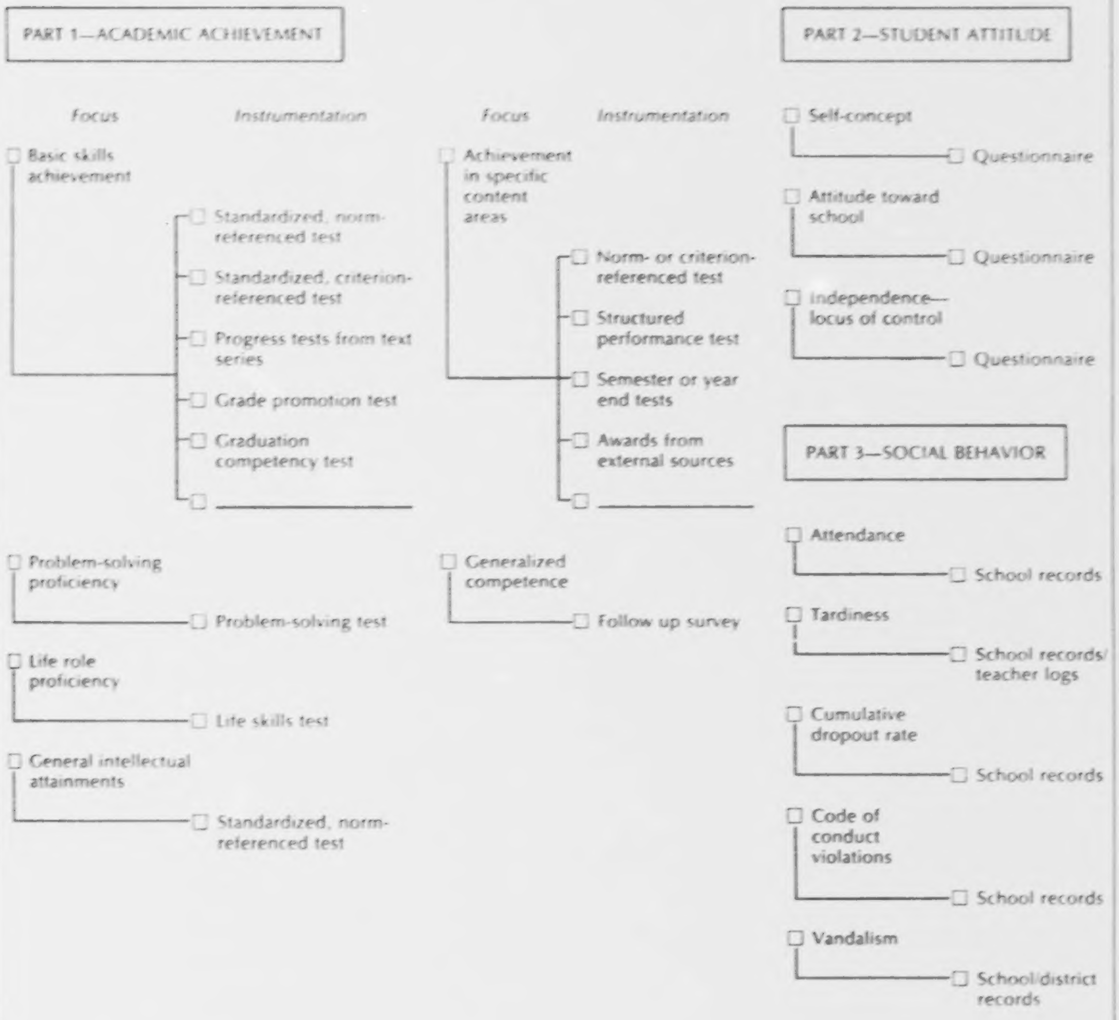


Figure 2. Profile Contents Menu.

PART 1—ACADEMIC ACHIEVEMENT

*Focus* — *Instrumentation*

- ☐ Basic skills achievement
  - ☐ Standardized, norm-referenced test
  - ☐ Standardized, criterion-referenced test
  - ☐ Progress tests from text series
  - ☐ Grade promotion test
  - ☐ Graduation competency test
  - ☐ ____________
- ☐ Problem-solving proficiency
  - ☐ Problem-solving test
- ☐ Life role proficiency
  - ☐ Life skills test
- ☐ General intellectual attainments
  - ☐ Standardized, norm-referenced test

*Focus* — *Instrumentation*

- ☐ Achievement in specific content areas
  - ☐ Norm- or criterion-referenced test
  - ☐ Structured performance test
  - ☐ Semester or year end tests
  - ☐ Awards from external sources
  - ☐ ____________
- ☐ Generalized competence
  - ☐ Follow up survey

PART 2—STUDENT ATTITUDE

- ☐ Self-concept
  - ☐ Questionnaire
- ☐ Attitude toward school
  - ☐ Questionnaire
- ☐ Independence—locus of control
  - ☐ Questionnaire

PART 3—SOCIAL BEHAVIOR

- ☐ Attendance
  - ☐ School records
- ☐ Tardiness
  - ☐ School records/teacher logs
- ☐ Cumulative dropout rate
  - ☐ School records
- ☐ Code of conduct violations
  - ☐ School records
- ☐ Vandalism
  - ☐ School/district records

## Preparing the Profile

Team members next make decisions about the appropriate display of data and who among them will complete individual displays. Each type of information is to be displayed on a single sheet of paper that can be used as a quick reference.

Once data have been summarized and displayed appropriately, team members prepare narrative descriptions of each summary, listing indicators of student performance represented by the data. Narratives are simple, declarative statements in plain language and avoid educational or testing jargon. Narratives describe but do not evaluate information contained in the data displayed and focus on key summary results, highlighting the most important aspects of student performance data. Team members write only a few statements that may be revised before the team finally settles on the exact wording. Data summaries may take one of several forms, including charts, graphs, and tables. Figure 3 is an example of data showing achievement in writing, and Figure 4 shows school attendance fluctuations.

Team members collect and organize these one-page data summaries into a single document, prepare a cover and an introduction, and print and distribute copies of the profile to all staff members for discussion. By covering several key areas, the team draws a current picture of student performance on a schoolwide basis.

## Using Data for Goal Setting

The completed profile becomes a tool for setting schoolwide improvement goals, as the leadership team directly involves the whole staff in generating a school improvement goal from the data already collected and compiled. The leadership team distributes copies of the profile to all staff members for consideration and evaluation in a final survey before making improvement goal decisions.

The profile evaluation process solicits feedback from all staff members. Staff members are asked to fill out individually the Profile Evaluation Worksheet (Figure 5) for compilation by team members as they prepare to set the school's improvement goal.

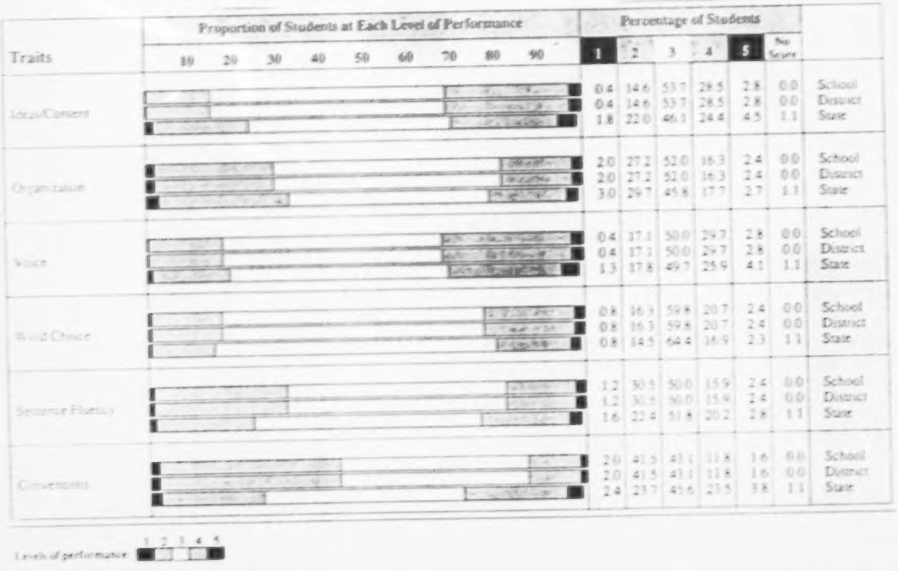
Staff members rate each narrative statement in all profile summaries on four bases:

1. Degree of *relative satisfaction* with the current picture of student performance, using a 10-point scale, from –5—very dissatisfied with result, nearly all students need improvement—to +5—very satisfied with result, nearly all students doing well in this area.

2. *Relative importance/priority for improvement* among areas with which staff members are dissatisfied, using a scale from 1, low priority, to 5, high priority.

3. *Acceptable standard of performance* in areas of least satisfaction, which would constitute "good" performance for students and result in a *high satisfac-*

**Figure 3. Writing Performance Assessment**

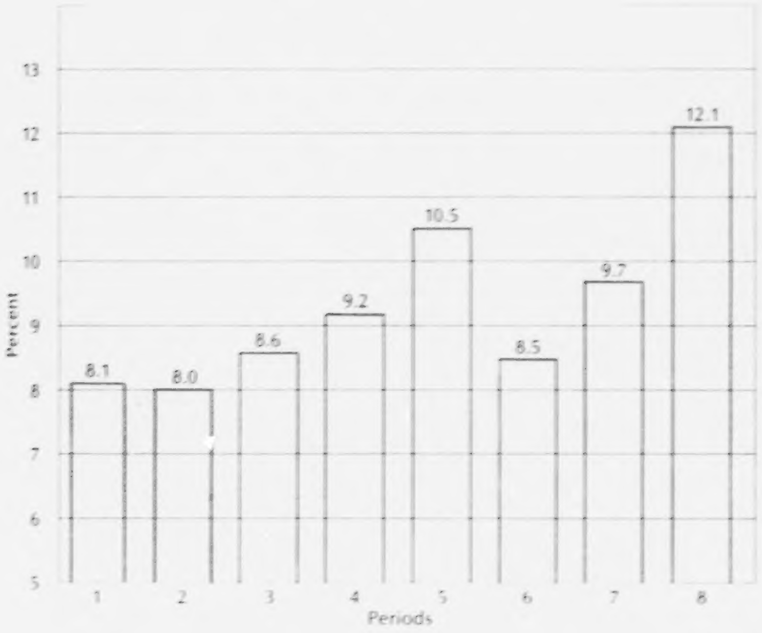


| Traits | | Percentage of Students 1 | 2 | 3 | 4 | 5 | No Score |
|---|---|---|---|---|---|---|---|
| Ideas/Content | School | 0.4 | 14.6 | 53.7 | 28.5 | 2.8 | 0.0 |
| | District | 0.4 | 14.6 | 53.7 | 28.5 | 2.8 | 0.0 |
| | State | 1.8 | 22.0 | 46.1 | 24.4 | 4.5 | 1.1 |
| Organization | School | 2.0 | 27.2 | 52.0 | 16.3 | 2.4 | 0.0 |
| | District | 2.0 | 27.2 | 52.0 | 16.3 | 2.4 | 0.0 |
| | State | 3.0 | 29.7 | 45.8 | 17.7 | 2.7 | 1.1 |
| Voice | School | 0.4 | 17.1 | 50.0 | 29.7 | 2.8 | 0.0 |
| | District | 0.4 | 17.1 | 50.0 | 29.7 | 2.8 | 0.0 |
| | State | 1.3 | 17.8 | 49.7 | 25.9 | 4.1 | 1.1 |
| Word Choice | School | 0.8 | 16.3 | 59.8 | 20.7 | 2.4 | 0.0 |
| | District | 0.8 | 16.3 | 59.8 | 20.7 | 2.4 | 0.0 |
| | State | 0.8 | 14.5 | 64.4 | 16.9 | 2.3 | 1.1 |
| Sentence Fluency | School | 1.2 | 30.5 | 50.0 | 15.9 | 2.4 | 0.0 |
| | District | 1.2 | 30.5 | 50.0 | 15.9 | 2.4 | 0.0 |
| | State | 1.6 | 22.4 | 51.8 | 20.2 | 2.8 | 1.1 |
| Conventions | School | 2.0 | 41.5 | 43.1 | 11.8 | 1.6 | 0.0 |
| | District | 2.0 | 41.5 | 43.1 | 11.8 | 1.6 | 0.0 |
| | State | 2.4 | 23.7 | 45.6 | 23.5 | 3.8 | 1.1 |

BEST COPY AVAILABLE

**Figure 4. Example of a Data Summary.**

Attendance per Class Period Grading Period 3



**Narratives**

1. The highest number of absentees occurs in the final class period of the day.
2. There is a higher percentage of absentees in periods directly before and after lunch than in morning class periods.
3. Overall, more students are absent in the afternoon than in the morning.

*tion* rating—a realistic, long-range standard to aim for in this area.

4. *A specific target for improvement*—staff members select from among all identified areas for improvement one or two short-term, time-based targets that are of *high priority*.

Using this feedback from the staff, the leadership team fills out a goal setting worksheet (Figure 6), restating narratives on problem areas as improvement goals and establishing standards of performance for those areas. After assessing priorities, the team selects one or two areas and sets a time-based target for improvement. They then report the improvement goals and targets to the staff, students, and community.

In this way, the profile becomes a tool for setting one or two improvement goals. The leadership team then works through the rest of the Onward to Excellence program to plan a prescription for improvement, prepare for and implement the prescription, monitor implementation, and evaluate progress toward reaching improvement goals. Underlying all decisions is a firm knowledge of baseline data about current student performance in the school.

Once the cycle of improvement is completed, schools can again profile the school, use new data to target another goal, and follow the process to plan for and monitor improvements. The cycle becomes a school-based method on regular collection and analysis of data on student performance.

**Figure 5. Example of a Profile Evaluation Worksheet.**

| AREA | | | | | | |
|---|---|---|---|---|---|---|
| Achievement | Behavior | Attitude | NARRATIVE | Satisfaction Rating | Importance Rating | Acceptable Standard |
| | | | Overall, 9 percent of students are absent from one class period each day. | | | |
| | | | The highest number of absentees occurs in the final class period of the day. | | | |
| | | | There is a higher percentage of absentees directly before and after lunch than in morning class periods. | | | |
| | | | Overall, more students are absent in the afternoon than in the morning. | | | |

**Figure 6. Goal Setting Worksheet.**

| AREA | | | | | PRIORITY | | | |
|---|---|---|---|---|---|---|---|---|
| Achievement | Behavior | Attitude | Goal and Target | Current Results and Standard | High Priority | Moderate Priority | Low Priority | Short-Term Target |
| | | | Improve daily student attendance. | Currently, 91 percent of students attend all courses; the standard is 97 percent. | | | | |
| | | | | | | | | |
| | | | | | | | | |

## Profile Checklist

The school profile provides baseline data about student performance against which effects of school improvements can be measured. This information becomes a primary tool in planning and managing targeted school improvement efforts.

There are several important factors to consider when creating a school profile:

- The profile describes student performance on a schoolwide basis.
- All students and all curriculum areas should be represented in the profile; the more comprehensive the profile, the more broad-based and complete the school picture that emerges.
- Creating a profile can sharpen the focus on both strengths and weaknesses in student performance.
- The profile is a snapshot of the school. It can become the first in a series that can be followed, perhaps annually, with updating and new improvement decisions.

Above all, the profile is a decision-making aid. Knowing specific information about student performance helps educators set goals and plan for improvements in student achievement, behavior, and attitude.

# VII-5

## Performance Assessment for Restructured Schools

Douglas E. Harris
National Gardening Association

Joyce Stone
Missisquoi Valley Union High School, Vermont

"Performance assessment is for probing and prodding the student's mind to reveal what it knows and what it can do in action." With these words, Grant Wiggins (1992) focuses our attention on the crucial question in our work to change schools and reform traditional curriculum. What do our students know and what are they able to do? How do we know that?

In this chapter we will attempt to answer these questions by linking assessment to the change process and to school improvement. We will examine the purposes of assessment with particular attention given to the implementation realities: supporting teachers and administrators, managing the logistical issues of time and resources, and developing effective communication about assessment within the school and the larger community.

In 1990, U.S. schools administered 127 million standardized tests at a cost of \$900 million dollars. What have decades of standardized tests yielded? For many educators, school boards, and communities they have not provided sufficient information about how their students apply what they have learned in real-life situations, how they make decisions and solve problems, what resources they are able to access in order to process and communicate what they have learned, or how they demonstrate their strengths and assess areas in which they need to grow. As we have sought to set national goals and to raise standards of achievement for all students, we began to "rethink" traditional practices, to revise curriculum, to change instruction, and to redefine what it means to be educated, capable, and prepared. Performance or alternative assessments have provided fuel for the process of change. Assessment has moved us forward and has forced us to look back.

In Vermont, the interest in authentic assessment has grown dramatically over the last five years. The initiatives at the state level have been influenced by and at the same time have made an impact on assessment programs in districts and schools throughout the state. All authentic or alternative assessments have in common their linking of a vision for education with research and practice, and the translation of these principles into curriculum and

---

Reprinted with permission from *Field Guide to Educational Renewal* by Holistic Education Press, P.O. Box 328, Brandton, VT 05733. 1-800-639-4122.

student outcomes. Once the learning outcomes have been identified, the standards and criteria for assessing are created and converted into portfolios, interviews, exhibitions, experiments, and other measures of performance. In this chapter, we will examine three case studies from the Franklin Northwest Supervisory Union in which local and state initiatives and national standards provide real-life examples of what works, what doesn't, and how assessment drives change.

## Assessing in Context

If decades of standardized testing at an investment of billions of dollars have not provided us the information we need to better serve teachers and learners, then it is time to change the tests. Grant Wiggins (1992) captures the strongest argument for authentic assessment in these words: "If tests determine what teachers actually teach and what students will study for—and they do—then the road to reform is a straight but steep one: test those capacities and habits we think are essential, and test them in context."

## Purposes and Misuses of Assessment

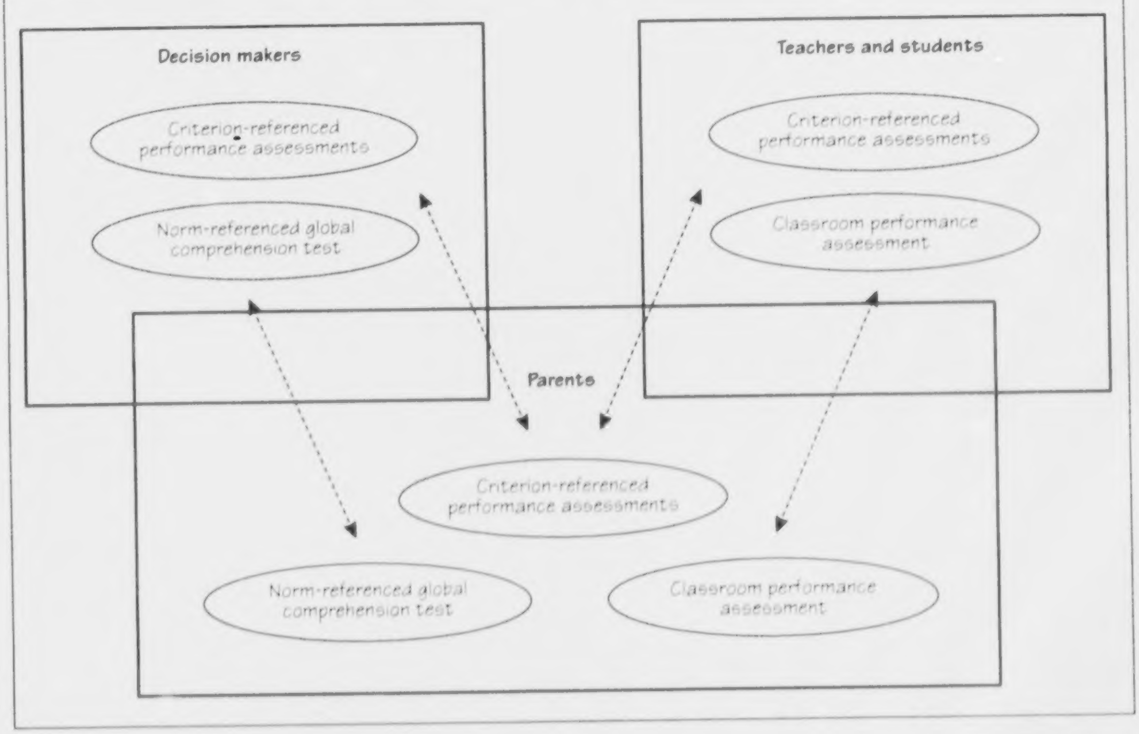
The purposes for assessment grow out of four fundamental processes which should be an on-going part of our schools: gathering information, focusing learning, making informed decisions, and revisiting the mission. At the heart of each one of the processes is the growing demand that schools be accountable to their communities.

Farr (1992) described three fundamental ways to gather information about student performance (Figure 1). These include (a) norm-referenced measures, which measure growth against an external group; (b) criterion-referenced performance assessment which measures growth against specified curricular outcomes; and (c) classroom performance assessment, which emerges from teacher/student interactions within the classroom.

Each of these types of information is valuable for certain purposes. Normative data is very useful for policy makers and is important to parents, but is frequently discounted as irrelevant to classroom decision making. Classroom performance assessment is the core of teacher decision making, provides imme-

Figure 1

THREE WAYS TO GATHER INFORMATION ON STUDENTS (FARR, 1992)


Decision makers
Criterion-referenced performance assessments
Norm-referenced global comprehension test
Teachers and students
Criterion-referenced performance assessments
Classroom performance assessment
Parents
Criterion-referenced performance assessments
Norm-referenced global comprehension test
Classroom performance assessment

BEST COPY AVAILABLE

diate feedback to students, and can be very interesting to parents, who rely heavily on the professional judgment of teachers.

Of the three, however, criterion-referenced performance assessment holds the most broad-based promise for educational change. This type of assessment forces clear outcome statements and serious discussion of ways in which to measure these outcomes. Inevitably, this process leads to discussion of teaching and learning. So long as parents, students, teachers, and policy makers are involved in specification of outcomes, and processes are established to ensure reliability and quality control, the information is valued by all participants in the educational community. Finally, information gathered related to outcomes forces reflection on the outcomes themselves, and thus can provide ongoing curriculum evolution.

Criterion-based performance assessment, then, serves all four purposes of assessment. It provides an efficient means of gathering information on student performance. To the degree that curriculum and instruction align with the outcomes, it provides "authentic" assessment; that is, assessment grounded in the teaching/learning process. It focuses learning, providing clear direction through the outcomes. It can lead to informed decision making as teachers, parents, and students learn to translate performance assessment into planning for future learning. Finally, it provides a powerful basis for reviewing the mission, so long as the mission and outcomes align.

Consequently, authentic assessment provides information about students in order to grade them fairly and evaluate their progress over time. It provides information to diagnose students' strengths and weaknesses. Schools may use assessment to motivate and focus learning by providing both teachers and students time for reflection.

The mathematics portfolio is an excellent example of how assessment may be used to make informed decisions. As a result of the national math standards on which Vermont's portfolio assessment is based, math educators have begun to revise curriculum and redefine math instruction. As more and more schools integrate math, science, and technology, these national performance standards will result in restructuring schoolwide. The fourth purpose for assessment, measuring how well we achieve our mission, is more elusive but in some regards the most crucial. Authentic assessment is tied to learner outcomes. If we meet those outcomes, assessment provides validation. If we do not, assessment is a powerful tool for change. In both cases, our commitment to public accountability demands that we measure where we are against where we said we were going—our vision.

There are also misuses of performance assessment. Criterion-referenced performance assessment should remain, at all times, criterion-based. Thus, any use of the measures to rank students, teachers, schools, or districts against one another (as opposed to reporting progress toward the measures themselves) must be avoided. If such ranking is essential for policy purposes, then normative measures should be employed.

A second misuse is direct application of outcomes-based results to teacher evaluation and reward structures. While student performance is the ultimate measure of educational success, and while teachers must be accountable for student progress, any use of outcome assessment for sanction or reward must occur in the context of considering other factors, including available resources, prior learning, competing socioeconomic and cultural factors. School, districts, and teachers must be accountable for outcomes, but this accountability must be contextual.

## Performance Assessment in Practice: A Model

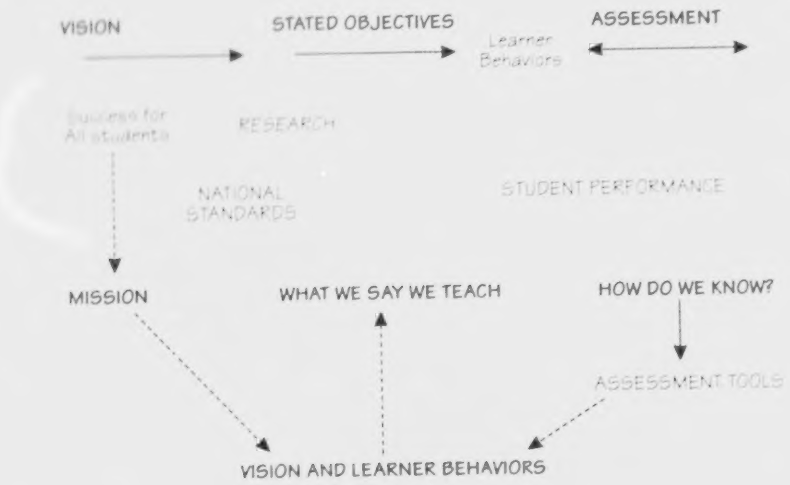
The model below (Figure 2) has been used successfully at the secondary level in the Franklin Northwest Supervisory Union to completely redefine the teaching and learning process in several content areas, among them language arts and foreign languages. Those working in assessment district-wide believe that the model may be adapted for use K-12 and that it will become part of a process of integrating content areas (departments) and restructuring the high school.

In fact, the model has been used at the elementary level. In 1990, Highgate Elementary teachers and their principal attended a School Development Institute for the purpose of developing a process to create a school vision/mission statement. Once that mission was in place, they used the Missisquoi Valley model to begin their process of curriculum revision and assessment. In that process, they asked the key questions embedded in the model. The answers gave the fuel to revise their implementation plan and to make critical decisions about teacher training, instructional strategies (such as cooperative group learning), and allocation of resources. The model has also been used to drive change at the district level, as implemented in the Franklin Northwest Supervisory Union.

The first step involves revisiting the vision. Do we have a vision? Is it compatible with the best thought in the field? Does it represent our local values? Is it still our vision? Have we been successful?

Figure 2

RENEWAL AND CHANGE MODEL AT MISSISQUOI VALLEY UNION HIGH SCHOOL



The next step requires that we research best practices in the field and identify appropriate external standards, if they exist. From this base, we revise or reinvent the curriculum. This process is two-fold: First, teachers must answer these questions: What and how are we going to teach? What are the learner outcomes? What are the knowledge, skills, and attitudes our students will possess at the end of the process? Assessment links teaching and learning directly to both the outcomes and the vision. Once teachers identify the why and what of teaching, authentic assessment answers the question" How do we know how well we and our students are doing?

Missisquoi Valley Union High School

"I've studied French for four years and I can't understand or talk to the tourists who come into my father's store."

"I want to work at the border, but I have to learn Spanish and pass a test. I took two years of French and don't remember anything. It'll take me the rest of my life to pass a test in Spanish."

"When I had Spanish in high school, all we did was conjugate irregular verbs."

Surprisingly enough, I have visited Montreal and Mexico City dozens of times and never, not once, did anyone ever ask me to conjugate an irregular verb, or a regular verb, for that matter. The three observations cited above happen to be the real words of residents in Highgate and Swanton, Vermont. However, the content of their meaning transcends town or state borders. These remarks are a serious indictment of the state of foreign language instruction in 20th century America. They tell us that the performance of our students does not match their needs in the context of life after graduation.

In 1986, three foreign language teachers, meeting after school in the French classroom, knew that they needed to define a new direction for the future. It was time to create a vision for the foreign language department that provided for success and competency of students in school and developed language proficiencies they could call upon in contexts outside of school. They agreed on a vision: proficiency and competency in all four language skills (reading, writing, listening, and speaking).

The next two years were hectic ones. We divided among ourselves a number of questions which had to be answered, What does out national organization ACTFL (American Council on Teaching Foreign Languages) have to say about language proficiency? Most of the teachers had heard of ACTFL Proficiency Guidelines (1980). The department head had been trained in the use of the guidelines and in oral proficiency interviewing. Where were "good things" happening in foreign languages? Distance permitting, we went to see the 'best practices" in foreign language teaching and discovered that good foreign language teachers were doing what good teachers in our own building were doing: student-centered/teacher-facilitated instruction, pairs and group work, hands-on activities, and instruction directed to all learning styles.

After gathering this information, we realized that the ACTFL Proficiency Guidelines contained the standards and measurement criteria which we would have to translate into our own curriculum scope and sequence. At the same time, we started to discuss assessment. How else would we know that our program was working? We needed to have feedback for ourselves, the school board, and the community—we were heading into uncharted waters. We were committed to several basic assumptions: that listening and speaking always precede reading and writing; that grammatical accuracy increases in a spiralling fashion and in the first years of language study it is far less important than the ability to communicate and be understood; that vocabulary should never be taught in isolation, but rather in a context which has some meaning to the student; that more is not always better; and that our mission year to year is not to complete a required number of chapters in the textbook.

By 1989, the Missisquoi Foreign Language Department had a proficiency scope and sequence in place, had identified student outcomes, and had piloted a number of new implementation strategies (conversational activities in pairs and groups and oral assessments based on proficiency standards in every unit). At the end of the first year, however, we were presented with a startling discovery: the textbook didn't work. The oral assessments conducted during and at the end of the year had led the teachers to believe that they were not going where they said they would, or at least not fast enough.

Teachers reported that they were frustrated with a text written to meet the old paradigm of foreign language teaching: grammar, vocabulary, and memorization. By that time, every teacher in the department had independently sought out training in teaching for proficiency. They were clearly well grounded and realized that a textbook published in 1986 and based on the pedagogy of the 1970s would not do the job. With the support of the central office, in the summer of 1990, they began to create their own thematic units to accomplish the department vision: competency and proficiency in the four language skills. At the same time, we developed a modified oral proficiency interview which would be administered at the end of the year to a random sampling of students in first year French and Spanish. The results of this assessment would provide concrete information on the success of our new curriculum and of our revised implementation strategy.

In the spring of 1991, 43 students (one in five) participated in the first round of modified oral proficiency interviews. In French the results exceeded the anticipated level of performance. In Spanish they were slightly below (which may be attributed to the fact that there was no Spanish program in the middle school at that time). In terms of the purposes for assessment identified earlier in this chapter, the interviews were exceedingly successful. They provided critical information which could be generalized into a program evaluation. This information could and would be used to make decisions about curriculum and instruction, restructuring the middle school program, and hiring additional teachers. For the teachers, the results provided a self-validation, a measure of the success of the mission of the department, and a source of professional and personal pride.

Since 1991, the assessment has been expanded to include random oral proficiency interviews at levels two and three as well. A portfolio program for writing begins in the second year of language study, is assessed and reported individually, and follows the student through his career in foreign language study. The teachers are currently identifying benchmark pieces for the writing portfolio assessment.

By any measure, the impact of this vision to assessment process has been a dramatic one on teaching and learning a foreign language at Missisquoi. Today the foreign language teachers have come to view learning on a continuum in which conceptual knowledge spirals over time into full control. As they work through that process, curriculum and student outcomes are based on practical, real-life situations which are found in authentic contexts. No one skill is taught in isolation, but rather all skills have an important role and are integrated into a whole. Competency in the language is no longer limited to structural (grammatical) competency. To the contrary, cultural (sociolinguistic) competency along with conversational competency are considered equally important. In fact, students are challenged to engage in linguistic problem solving by putting all their competencies to work in finding their way out of a role play situation for which they have never received direct instruction. This is the broadest standard for competency.

Let's imagine that we return in 1993 to the former student in Highgate whose father owns a store near the interstate. He expressed regret and frustration about not understanding or being understood in French. Today, would he be able to help a visitor from Montreal who enters the store on Sunday afternoon and points to a flat tire? The man makes it clear that his spare tire has been stolen. Our former student must tell the man that, because it is Sunday, no garage in Swanton is open. He will have to go south to Saint Albans and travel three miles off the interstate to find the only garage with tires for sale and an attendant to help him. It is quite likely that this student, with his four years of French, will be able to do the linguistic problem solving to work his way around the words he doesn't know and make himself understood. It is absolutely certain that in his French class he has practiced French in many situations similar to this one and has done so successfully.

### Highgate Elementary School

Highgate Elementary School, a K–6 school of about 400 students, serves a rural dairy fanning community nestled on the Canadian border in northwestern Vermont. Highgate is one of Vermont's poorest communities, with both property values and per capita income substantially below the state average.

In 1990, Principal Brian Duprat began his tenure at the school by involving the community, staff, and school board in developing a vision statement for the school. This statement has evolved into the basis for all operational decision making within the school. Since no explicit vision statement preceded this one in Highgate, the focus within the school was forward-looking rather retrospective. The staff and board spent much time that first summer defining ways to reach the vision, but little time in reflecting on where the school had been.

At this same time, the State of Vermont was beginning implementation of a portfolio assessment process in grades four and eight. Because the portfolio assessment in writing and math is grounded in research (process writing and the 1988 Standards of the National Council of Teachers of Mathematics, respectively) and because of the presence of some very strong teachers at grade four, implementation of the statewide portfolio became a natural vehicle for beginning forays into outcome-based criterion-referenced assessment.

Highgate's grade four teachers participated in statewide training as did teachers throughout Vermont. Beyond that, however, they engaged in dialogue with their principal, with central office staff, and with teachers throughout the district in comparing the written and taught curriculum to the standards.

The teachers also joined colleagues throughout the state in developing rubrics to articulate performance standards for math and writing.

Review indicated that, while implementation of the writing process was somewhat uneven, the process was being implemented in all classrooms. Thus, the challenge was to examine existing practice, to rethink ways to communicate with students and parents about writing, and to take on the unfamiliar task of applying rubrics to student writing.

The task in mathematics was much more formidable. As we examined the curriculum against the standards, we found major gaps, particularly in the areas of estimation and statistics/probability. Further, we found that lack of experience in assessing the language of mathematics and in mathematical representations made it difficult to apply the rubrics to even those areas of the curriculum in which our practices matched the standards.

The fourth grade teachers worked through this process in terms of the requirements of state assessment. While the initial reporting at the state level was not as reliable as it could have been, the teachers found that the assessment process profoundly impacted the way they viewed children as writers and mathematicians. In addition, the portfolio process yielded a common vocabulary to discuss teaching and learning in all curriculum areas.

Meanwhile, the forced focus at grade four led to review of the entire writing and math curriculum. This in turn led to consideration of materials and supplies for teaching, particularly in mathematics. Further, the grade four teachers have become resources for their colleagues as we have expanded authentic assessment into grades five and six.

There is much yet to do. We need to refine our reporting processes to make certain that useful information gets to students, parents, and to one another. We need to find ways to extend authentic assessment into our primary unit, and to expand the process into all subject areas. We need to continually revise curriculum and instructional practice, based on outcome data.

Thus far, assessment has begun to redefine Highgate as a school in which reflective practice is based on student outcomes and upon careful dialogue among educators, students, and parents. As we work to get better, assessment will remain a powerful tool and an anchor to measure our performance.

### Franklin Northwest Supervisory Union

Both Highgate Elementary School and Missisquoi Valley Union High School are part of the Franklin Northwest Supervisory Union. The district consists of four elementary schools, a middle school, and a high school and serves approximately 2600 students.

Franklin Northwest has made a conscious commitment to criterion-based performance assessment as its main vehicle for school improvement. This commitment has translated into district-wide goal setting, in revision of curriculum, in staff supervision and evaluation, and in allocation of resources. Planning at all levels (classroom, team, building, administration, and board) reflects this shared commitment.

For each of the past three years, increased accountability through assessment of student performance has been a stated goal of the supervisory union school board. This goal has translated into action plans as follows:

- 1990–1991: Review and revision of curriculum and beginning of implementation of performance-based assessment in writing and math, grades 4 and 8.

- 1991–1992: Performance assessment as a major staff development focus; continued implementation of performance assessment in writing and math, grades 4 and 8.
- 1992–1993: Specification of content-based outcomes, benchmarked in grades 8 and 12, for math English, science, social studies; implementation of performance assessment in writing and math, grades 4 through 8.

The board goals are incorporated into goals at the school level and considered in collaborative planning at all levels. Progress toward district- and schoolwide goals are reported by each building principal.

Each year, we review progress towards these goals at the school and district level. We also are building student performance outcome data for program evaluation. While this is in its infancy, it ultimately holds the greatest potential to moving the district.

The district has made assessment its primary priority in allocating district-level staff development resources. Franklin Northwest's participation in ASCD's Consortium on Authentic Assessment has paid tremendous dividends as we have worked directly with districts in the U.S. and Canada and have benefited by sharing best practices and dilemmas. The supervisory union has also developed local training in performance assessments through symposia and release time for teachers to work in this complex area. Schools and individual teachers have likewise set priorities in this area.

A strong "yes" to resource allocation in this area means "no" in other areas in a poor district. This forces teachers and principals to carefully evaluate how time and resources are allocated. While this can cause consternation in a time of competing opportunities for improvement, it has also focused dialogue and made a powerful statement about what we consider most essential.

Teachers' use of assessments of student performance is a major component of our staff evaluation system. In our clinical supervision model, principals and teachers discuss assessment as a major part of our preconference. Principals specifically include assessment in the observation process, both in terms of formal assessment of criterion-based outcomes and classroom performance assessment. The post-conference includes discussion of ways in which the assessment data will impact instruction, and ways in which the information will be shared with parents.

One of the most important factors in focusing on assessment has been our resisting the temptation to move on. While this has been a major priority for three years, we know that it will take at least three more years to complete the process of developing outcomes and criterion-based assessment processes across the curriculum and grade levels. This long-term commitment tries the patience of all of us, particularly when everyone recognizes the importance of getting to where we want to go. Nevertheless, the fact that we focus rather than shotgun our resources has done a great deal to unify the district while honoring differences of implementation strategies within the schools.

## Conclusions: Implementation Realities

In Franklin Northwest and in our individual schools, we have come to grips with certain implementation realities. For the sake of discussion, we will group these in two categories: non-negotiables and support strategies.

### Non-Negotiables

In developing our commitment to criterion-based performance assessment, we have made a strong commitment to collaboration. Teachers have major input and frequently autonomy in developing implementation strategies. Principals and their staffs have much discretion in setting timelines and selecting materials and methodologies. Nevertheless, an implementation reality is that certain non-negotiables exist within our district. Once these are recognized and understood, it becomes much easier to focus our energies.

*Non-negotiable 1.* Results will be used for accountability and program improvement. Regardless of specific methodologies, we will be accountable for student performance. This includes a commitment to measure performance, to ponder its implications, and to adjust our program accordingly.

*Non-negotiable 2.* The information will be communicated to all relevant parties. Results belong equally to students, parents, educators, and policy makers. All information, once gathered, will be shared, whether the news is good or bad.

*Non-negotiable 3.* The information will actually be used in decision making. Beyond sharing the information, we will ensure that it impacts the curriculum, instructional practice, and resource allocation.

*Non-negotiable 4.* All our work in assessment will embrace these values: a belief in student self-evaluation, a commitment to active learning, and a commitment to collaborative practice.

## Support Strategies and Survival Techniques

In supporting teachers in implementation of criterion-based performance assessment, we must provide educational, logistical, and moral support.

Educationally, we must give teachers the opportunities to learn about criterion-based performance assessment. This includes not only the theoretical basis but also practical experience in implementation.

Within our district, we have provided coursework, symposia, and in-service activities. We have also encouraged staff members to participate in regional and state training and have provided release time for this to occur. We have built a substantial library of print and nonprint resources on performance assessment and have worked very hard to encourage staff members to pursue coursework in this area.

Perhaps most importantly, we have encouraged teachers to team in performance assessment and have provided release time for team planning.

Logistically, we have provided release time for development of outcomes and for implementation of performance assessment practices. We have drastically reduced normative testing (from 13 subtests to 2) and have worked to minimize distractions as assessments occur.

As we define our survival strategies, we find it absolutely necessary to improve the logistics of dealing with the volume of paper generated by performance assessment and to explore ways to use technology to aid in storage and retrieval. Much work remains in this area.

Undoubtedly, the most challenging area is moral support. As teachers learn new techniques and explore student outcomes, they are constantly confronting the unknown and unfamiliar. As in any learning situation, teachers, as learners, must know that mistakes are not only all right, but essential. Likewise, they must know that disagreement is essential, that conflict will occur, and that sacred cows (some of which belong to them) will be slain.

We try to provide unqualified support, within the parameters of the non-negotiables listed above. We also work to celebrate diversity, to publicize our successes, and to see our failures as opportunity.

Nevertheless, we frequently get overwhelmed. When that occurs, we return to the process outlined in the case studies. We revisit our vision, check our practice against research, affirm our premises, and move on.

We in Franklin Northwest are convinced that all school change is ultimately student-centered. We believe, as our high school principal often says, that "you have to draw the target on the wall." By drawing the target through specifying outcomes, by agreeing on means to assess these outcomes, and by sharing our results, we believe we can make our schools better places for children.

Our work is at its infancy; yet, the dividends are already obvious. If we continue to focus and to challenge ourselves in this area, they will multiply.

## REFERENCES

Farr, R. (1992). Putting it all together: Solving the reading assessment puzzle. *Reading Teacher*, 46(1), 32.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

Vermont Department of Education. (1991). *Vermont mathematics portfolio resource book*. Montpelier, VT: Author.

Wiggins, C. (1992, May). Creating tests worth taking. *Educational Leadership*.

# VII-6

## Assessing Student Performance in Crater High School's Applied Technology Program

NANCEY OLSON
NORTHWEST REGIONAL EDUCATIONAL LABORATORY

Crater High School is a comprehensive high school in the Central Point School District in south-central Oregon. With the assistance of several grants from the Oregon Department of Education, Crater High School has undertaken several projects to improve the educational programs for its more than 1,200 students, including the restructuring of the school's Applied Technology program. Whereas traditional instruction and assessment focus on teaching the facts and figures of a subject and asking students to recall that information, technology education focuses on the "application of human knowledge." The restructuring of the technology classes focused on providing opportunities for students to reach into the application and synthesis levels of learning.

The vision statement for the Vocational Technology Department of Crater High School states that staff members are

> . . . dedicated to providing a total dynamic educational program that helps students prepare for an ever changing technological society.
>
> We aspire to excellence through a practical delivery system that encourages critical thinking, problem solving, self-esteem, skill mastery, and life-long learning; utilizes appropriate technologies; and integrates curricular areas. Working with industry and the community, technology teachers facilitate student centered activities in an Applied Technology Learning Center serving all students.

The Applied Technology Learning Center was designed around self-contained work stations, termed "modules," that support self-directed, individualized instruction. As planned, the modules contain everything needed to complete an assigned work

---

This chapter is adapted from N. Olson, *Curriculum Aligned Products: Applied Technology Program*. Portland, Oreg.: Northwest Regional Educational Laboratory, November 1993).

task. The modular concept provides flexibility and cost-effectiveness because many topic areas can be addressed without the expense of equipment duplication and excessive lab space.

## State Learning Goals and Curriculum Alignment

Schools in the state of Oregon have been charged with preparing students for life in the 21st century. One indication of the state's commitment to this goal is the development of the Certificate of Initial Mastery (CIM). The CIM is intended to both define the desired outcomes and to establish criteria to measure student performance related to the outcomes. School sites across Oregon are involved in pilot projects that will serve as models for other districts as they move to meet the state requirements for implementing the CIM. (See Section III of this guide for a draft of the CIM outcomes and extended outcome definitions.)

Crater High School was one of several schools chosen to receive a School Performance Assessment Network (SPAN) grant. The school used this one-year grant to explore alternative assessment strategies related to the required CIM.

During this year of experimentation, district leaders also chose to work with staff and community to develop a district mission statement and student learning goals that would align with the state learning goals. Teachers involved in the development of alternative strategies for assessment also wanted their course or classroom goals to align with the state and district goals so that their assessment strategies would reflect all three sets of goals.

### CIM Outcomes

The application for the SPAN grant indicated that Crater High School would assess the following CIM outcomes:

- Demonstrate the ability to work individually and in groups to define problems, develop and evaluate solutions, and apply strategies for solving life problems.
- Demonstrate the ability to communicate effectively in a variety of modes.
- Demonstrate an understanding of scientific and mathematical concepts and processes and technological advances and how they affect the world.

These CIM goals were taken from a draft document from the Oregon Department of Education and were part of the grant application. They have since been edited and refined, but still reflect the essence of the final set of CIM outcomes approved by the state.

### District Goals

The district goals listed below are arranged in order of priority as approved by the Central Point School District Board of Education.

#### Mission Statement

School District #6, in partnership with parents and community, is dedicated to assisting students reach their individual potential as productive, responsible participants in society. Within an environment that encourages learning, students will acquire skills, knowledge, and attitudes necessary to successfully meet the opportunities and challenges of the 21st Century.

#### Goals

Students completing the District #6 educational program and entering the adult world shall have:

** proficiency in the basic skills
** mastered higher level thinking skills
** the ability to cooperate and collaborate effectively within a group
* integrity, ethics, respect and responsibility for self and others
* a sense of self-worth with an understanding and commitment to a healthy life-style
** the curiosity and skills for life-long learning
* awareness of historical roots, global responsibility, and appreciation of cultural diversity
** acquired career readiness skills and an appropriate work ethic
** practical experience in the area of technology
* appreciation for literature, fine arts, and the creative process
* exposure to and opportunity for mastering foreign language.

The goals marked ** show a relationship between CIM outcomes chosen for the SPAN grant and district goals. The assessment strategies developed by the teachers and students in the technology classes were also designed to reflect the students' accomplishment of these goals.

### Technology Systems Goals

The course goals for the first semester class, Applied Technology 101: Tools of Our Time, were as follows:

1. Demonstrate the ability to direct one's own learning, think critically and reflectively.
2. Establish an awareness of technology and its relationship to change and natural resources.
3. Gain an understanding of problem solving techniques and their real world application.

4. Deal with information as a human resource and common technologies for information management.

In addition, the thinking skills that were specifically addressed through the class activities included these:

1. *Application.* Student will use learned material in new and concrete situations.
2. *Analysis.* Students will research, interpret, and break down material into component parts so it will be more easily understood.
3. *Synthesis.* Students will put material together to form a new whole in the context of a multimedia presentation.
4. *Evaluation.* Students will judge the value of material and organize material in a portfolio.

For two semesters, the technology classes devoted their energies to implementing a new design for both teaching and assessing student performance in the identified goal areas. Students were involved in the development of the scoring procedures, and the teaching team decided to have students collect their work in portfolios, recording the students' performance on assigned tasks as they progressed through the work stations and evaluating projects required for the course.

## Task Description

The pilot of the new delivery system in the applied technology courses entitled Tools of Our Time and Advanced Tools of Our Time enrolled 311 students for two semesters—81 seniors, 62 juniors, 83 sophomores, and 85 freshmen. Students were required to complete all assignments to receive a grade. Students not completing all assignments received an "I" for In Progress, and were given the designated time allowed by Crater High School guidelines to finish the work. Grades were translated to traditional A, B, C, D, F for semester report cards.

### Course Assignments

In the class called Tools of Our Time, students worked in teams of two or three that rotated through the computer work stations to complete the following learning units, including all tasks as designated on the assignment sheet:

1. Superpaint (drawing and painting)
2. Animation Works
3. Interactive Laserdisc
4. CD-ROM
5. Publish It Easy (desk-top publishing)
6. Xapshot (still camera)
7. Videospigot (video digitizer)
8. QuickTime (video editing)
9. Modem and Satellite (distance learning)
10. Optical scanner (importing pictures)

A sample assignment sheet appears in Figure 1.

FIGURE 1

ASSIGNMENT SHEET

Name: ______________________

Date: __________ Period __________

DIRECTIONS: Complete each of the following assignments. Have your instructor initial each one upon completion. Keep this portfolio as a record of your work.

1. ___ Create a folder on your floppy to save your files (use a nickname)
2. ___ Complete the Macintosh Basics Tour
3. ___ Complete the Microsoft Works tutorial
4. ___ Complete the Microsoft Works tour
5. ___ Complete the hypercard practice stack
   ___ drawings ___ animation
   ___ buttons ___ sound
   ___ text
6. ___ Complete a story board for a hypercard stack (topic of your choice)
7. ___ Create a multimedia report containing the following (Use the story board for a guide)
   A. Hypercard (practice stack parts)
   B. Sound
   C. An imported xapshot
   D. An imported scanned picture
   E. An imported CD/ROM picture
   F. Clips from an interactive laser disk
   G. Desk-top publisher or Microsoft Works or both
8. ___ Participate in team exercises
9. ___ Participate in team evaluation (network groups)
10. ___ Develop and organize a portfolio
11. ___ Keep a daily log of your activities and attendance

### Final Evaluation

Each student also had to complete a project as part of his or her final evaluation. The purpose of the project was to integrate the skills learned at each of the module centers. The students helped develop the evaluation criteria, and each of the projects was evaluated by the teachers and a team of students. The final evaluation task was given to the students at the beginning of the semester and required that each student demonstrate the skills acquired at each of the

BEST COPY AVAILABLE

15 work stations. A sample Final Evaluation Sheet appears in Figure 2.

In addition, all students were to keep a portfolio of their work. This notebook included their work to date and a daily log of their learning activities. The teacher checked the portfolio every three weeks and graded it using a Portfolio Rubric.

### FIGURE 2
### FINAL EVALUATION

Project Content:

**One** stack that shows what you've learned and done at each station. Not a separate stack for each station. The stack will include the following items.

Station:

1. Your own recorded sound and sound copied from "cartoon."
2. Animation works tutorial saved to your disk but not in your project.
3. Beethoven not needed in your project.
   A laserdisc "clip list" saved to your disk.
   Hypercard buttons to play laserdisc if time allows.
4. Publish It tutorial saved to your disk and printed.
5. Xapshot pictures imported into project.
6. Desert storm not needed in your project. Same as Station 3.
7. Video spigot saved to your folder on the shared disk.
8. Clip list.
9. World Atlas information not required, extra credit for maps or information from world atlas put into your project.
10. Clip list.
11. Same as Station 1.
12. Picture, map, and information from Grolier Encyclopedia put into your project.
13. Scanned picture put into your project.
14. Time table information not needed in your project.
15. Clip list.

Animation Works and Desk-top publishing should be saved to your disk (or the shared folder) but not included in your project.

You need at least one clip list from a laserdisc saved to your disk. You may receive a sheet later for making a button to play the laserdisc.

In addition you need: a title card; author card; icon buttons; visual effect on buttons; hypercard animation clip art; fields; original drawings.

Extra credit will be given for:

1. Unique and creative use of equipment and programs.
2. A "theme" to your project.
3. A project that could be used for credit in another class.
4. A project that can be used to teach something.

## Data Displays

Every student in the technology classes received individual feedback on his or her progress on the assigned tasks and the final evaluation. Both teachers and teams of students evaluated students' performance. The scoring rubrics used in this evaluation, however, provide an incomplete picture, telling us little about how students' skill levels were assessed. The rubrics don't tell us that students demonstrated their level of skill through the presentation of projects. The presentations reflected the students' ability to integrate the skills and to produce a product reflecting the various technologies and the students' creativity. When students made their presentations, they had to use the computer skills they learned; it would have been nearly impossible for them to adequately demonstrate their learning via a paper-and-pencil test. A Presentation Skills Rubric helped us assess this component of student performance.

We also used portfolios to collect each student's work, teacher and group evaluations, self-reflections, and a daily log of the student's work. Included in students' portfolios were their responses to a Work Habits Rubric and a Self-Assessment Inventory administered every three weeks. These assessments show how students view themselves as contributing members of a learning or work environment. A Portfolio Rubric was also used to track students' maintenance of the portfolio and completion of assigned tasks.

### Simulation Data

The teachers of the Tools of Our Time courses did not aggregate the students' scores to produce data showing the performance of the class as a whole. For the purposes of this chapter, we reviewed the scoring rubrics and created *simulation data* to reflect the type of data that could be presented if desired. The main reason for aggregating data would to be to assess overall student performance on the assigned tasks to see patterns of strengths and weaknesses. An analysis of the results would lead to discussions about program revisions or changes in instructional practices when needed.

Figures 3 through 9 illustrate what a complete data display would look like if it included in a school profile. They present examples of aggregated data that could be generated from the Work Habits Rubric, the Presentation Skills Rubric, and the Portfolio Rubric. The data displays have three parts: (1) a prologue to explain how the data were collected, (2) a graph to illustrate the data, and (3) narrative statements. The graphs show only one year of *hypothetical* data, but they could include one to five years of data if updated from year to year.

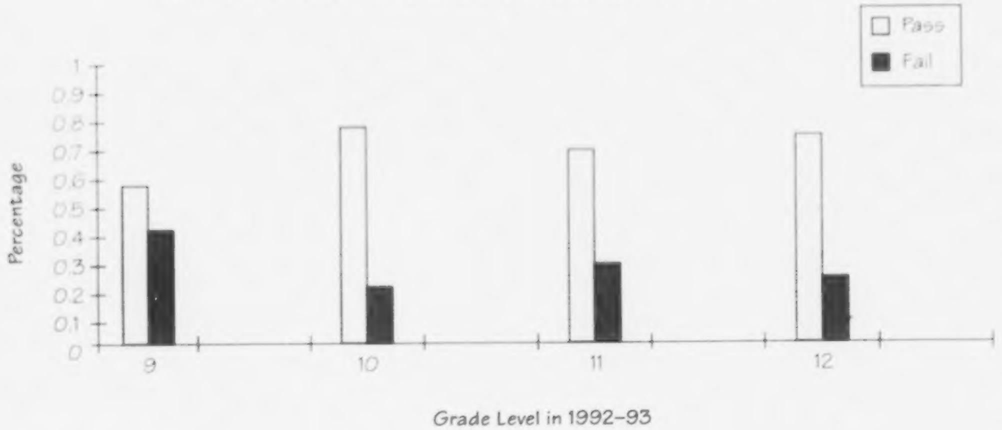
BEST COPY AVAILABLE

FIGURE 3

PASS/FAIL RATES ON CIM OUTCOMES BY GRADE LEVEL

**Prologue**

Students were evaluated on CIM outcomes using rubrics that rated students' work habits, portfolios, and presentations from 1 (low) to 5 (high). Teachers later translated these performance ratings into traditional letter grades for official reports at the end of the semester.



**Narratives**

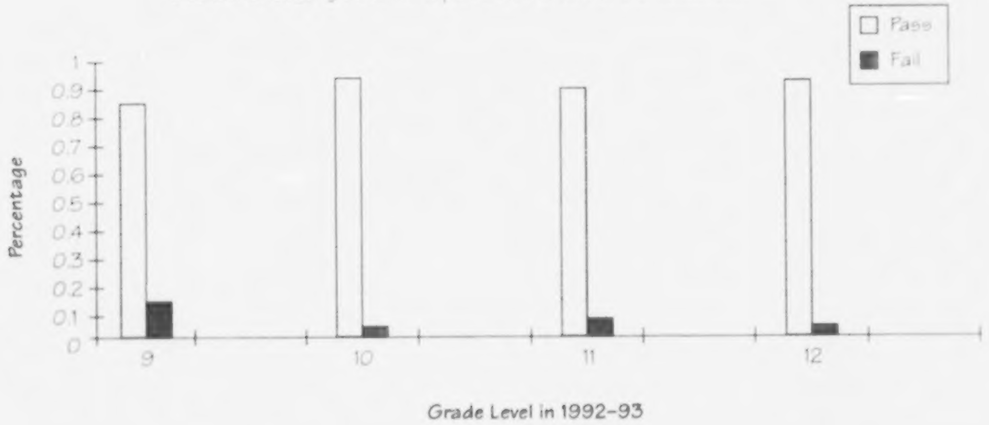
1. Tenth graders have the highest passing rate at 78 percent; 12th graders are next at 75 percent.
2. Ninth graders have the highest failing rate at 42 percent.

FIGURE 4

PASS/FAIL RATES USING TRADITIONAL GRADING BY GRADE LEVEL

**Prologue**

Students were evaluated on CIM outcomes using rubrics that rated students' work habits, portfolios, and presentation skills from 1 (low) to 5 (high). Teachers later translated these performance ratings to traditional letter grades for reports at the end of the semester.



**Narratives**

1. With traditional grading, 10th graders (94 percent), 11th graders (91 percent), and 12th graders (93 percent) had the highest passing rates.
2. Ninth graders had the lowest passing rate (85 percent) with traditional grading.

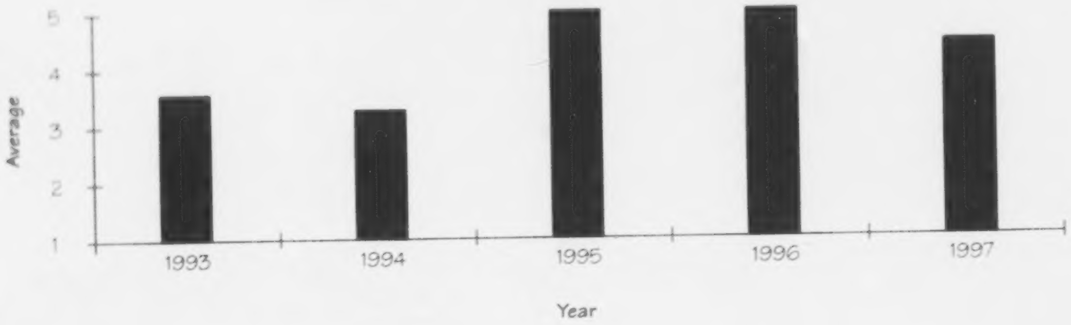
BEST COPY AVAILABLE

FIGURE 5

## WORK HABITS ASSESSMENT: ALL SKILLS, 1993–1997

**Prologue**

The Work Habits Rubric evaluated students on a scale from 1 (low) to 5 (high) in seven different categories: Attendance, Quality of Work, Dependability, Knowledge of Job, Quantity of Work, Teamwork, and Initiative. All scores across all categories were calculated at the end of the semester to obtain an average score for all students in all classes.



**Comments**

There are several options for displaying the data from the Work Habits Rubric. All skills and traits can be averaged and charted as above, or each trait can be tracked separately as illustrated in Figure 6. In addition, the averages for each trait could be compared for one year, as shown in Figure 5.

FIGURE 6

## WORK HABITS ASSESSMENT: TEAMWORK

**Prologue**

The Work Habits Rubric evaluated students on a scale from 1 (low) to 5 (high) in seven different categories: Attendance, Quality of Work, Dependability, Knowledge of Job, Quantity of Work, Teamwork, and Initiative. This chart shows the average score in the Teamwork category, calculated at the end of the semester for the last evaluation.

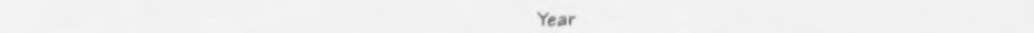


FIGURE 7

### WORK HABITS ASSESSMENT: AVERAGE SCORE PER CATEGORY AT END OF SEMESTER, 1993

**Prologue**
The Work Habits Rubric evaluated students on a scale from 1 (low) to 5 (high) in seven different categories. The average score in each category was calculated at the end of the semester for the last evaluation.

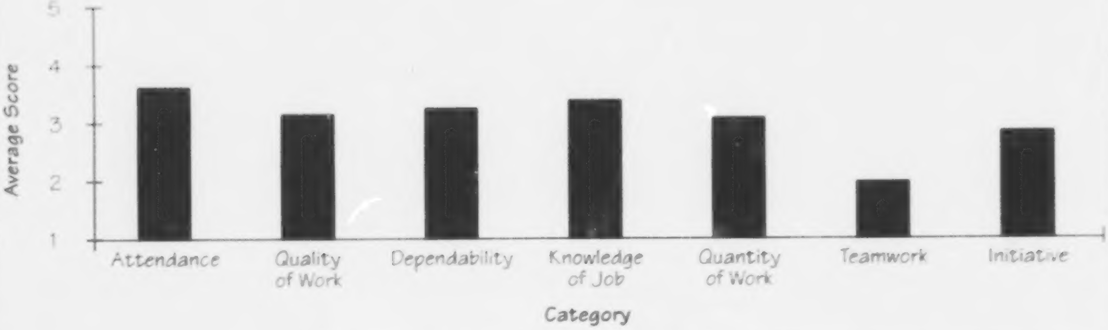
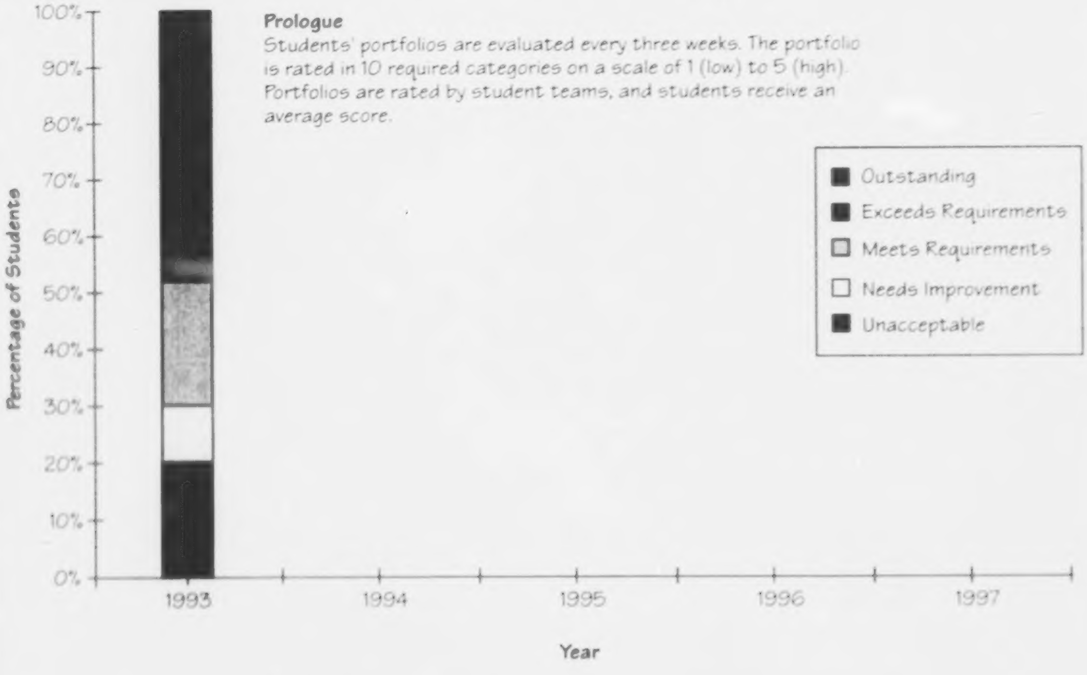


FIGURE 8

### PORTFOLIO EVALUATIONS: PERCENT OF STUDENTS IN EACH OF FIVE LEVELS AT END OF FIRST SEMESTER



**Prologue**
Students' portfolios are evaluated every three weeks. The portfolio is rated in 10 required categories on a scale of 1 (low) to 5 (high). Portfolios are rated by student teams, and students receive an average score.

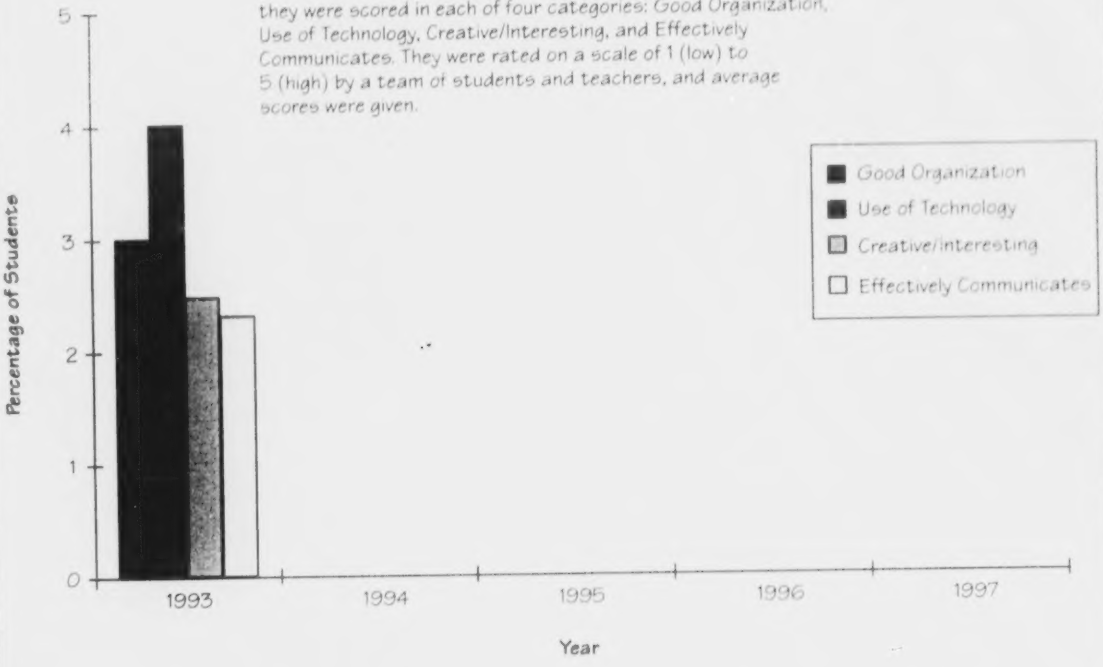
**Comments**
Data would be collected over a five-year period, and narrative statements would reflect trends over time.

BEST COPY AVAILABLE

FIGURE 9

STUDENT AVERAGE BY PRESENTATION SKILL AREA
(FINAL EVALUATION AT END OF SEMESTER

**Prologue**
Each time students made a presentation of an assignment, they were scored in each of four categories: Good Organization, Use of Technology, Creative/Interesting, and Effectively Communicates. They were rated on a scale of 1 (low) to 5 (high) by a team of students and teachers, and average scores were given.



**Comments**
The Presentation Skills Rubric lends itself to presenting average scores similar to those in Figure 3. The example shown here gives additional information indicating students' areas of strengths and weaknesses in the skill categories.

## Summary: A Report*

### Overview

The students who participated in the Applied Technology program were informed of the nature of the restructuring project and related CIM outcomes. Students enjoyed the opportunity to develop evaluation tools and took an active part in this experience. Instructors found the initial stages of implementation challenging, but once the basic structure of the program was in place, course management was well within the comfort range of both instructors.

The teaching staff developed the curriculum as a team, with the computer science teacher handling most of the equipment set-up and the technology teacher handling most of the lesson plans and information dissemination.

### Positive Experiences

Placing students in a professional atmosphere created a learning environment relatively free of discipline problems. We also found that absentee rates in the test group were quite a bit lower (5.4 percent) than district averages (15 percent reported

---

*All of the comments in this summary are taken from a report by technology instructor Dave Miller and computer science instructor Dale Sandberg. They presented their report to the SPAN committee, which was responsible for coordinating district efforts to create alternative assessments for the outcomes mentioned in the grant.

December 14, 1992). The quality and sophistication of students' work increased as all involved raised their expectations and as instructors grew more adept at working with the new delivery system. Very few students dropped the course, even though expectations were considerably higher than in traditional shop classes, and very few students failed to complete all the assignments (three to seven per class).

As our organization of the program improves, the opportunities for instruction increase. The system provides ample opportunity for collaborative learning and individual work without tracking students by ability or grade. It is competitive only if students wish to make it so. In addition, many students can effectively learn with a minimum number of instructors (30 to 1 works quite nicely).

Virtually any subject can be integrated into the course without much difficulty if other department members are willing to work with us. Our curriculum integration could be improved if we were not confined to Carnegie units and seat time.

Demonstrations of knowledge eliminate the possibility of cheating or copying found in traditional pencil-and-paper tests. They also provide students a wide variety of methods and topics through which to demonstrate their abilities. Giving students more opportunities to evaluate courses and select topics resulted in more time on task and increased the quality of their work. And the evaluation of work habits helped students develop a realistic understanding of the strengths and weaknesses of their own work ethic.

### Problems and Areas Needing Improvement

**Portfolios** provided an excellent evaluation tool for instructors, but students had a difficult time understanding their value. Because students didn't use the portfolios to study for tests or to collect factual notes as in a traditional class, they viewed them as busy work. It was difficult to maintain regular input and upkeep. The physical management of portfolios was also difficult: trying to store 300 portfolios while maintaining student access was almost impossible. When students were asked to keep track of the portfolios, the portfolios were often lost, damaged, left in lockers, and so on. We experimented with electronic portfolios and believe this approach may hold promise. If students have one portfolio for their entire program rather than a separate one for each course, management may be easier. Evaluation might be done by the "Tech. Prep." committee or some similar group.

The initial set-up of **equipment and software** requires funding for both preparation time and the purchase of equipment and software. Once in place, however, this delivery system is extremely economical. Continued growth of this project will require periodic updating of equipment and software. Although grant monies got us up and running, it is essential that funding be maintained if the Tech. Prep. program is to be completed.

**Current report cards** do not accurately reflect outcomes. Some consideration should be given to change in reporting methods. Thought also should be given to the **awarding of credit to students.** For instance, students selecting an "academic" topic for demonstrating technology skills should be able to receive credit in both the academic course and the technology course upon demonstration of skill and knowledge. Allowing students to receive two credits for one time period would empower students in their learning, ease scheduling problems by providing more opportunity for acquiring necessary credits, and provide a strong nucleus for Tech. Prep. English, Math, and Science.

The **completion of courses** is also a concern. Students receiving an "In Progress" rating on their report card do not have adequate provisions in the school day to complete work. Do we keep these students in the program next semester until they've caught up, form a new "make-up" class, fail students after a specified period of time, or try some other solution?

The students involved in this pilot were chosen at random from our traditional program with traditional experiences. During the first semester, students and teachers both made significant adjustments to the new learning environment. The results may not accurately reflect how students might perform if they had come through the restructured system beginning in elementary school. We suspect that students already familiar with rubrics, portfolios, and self-directed learning environments will do better.

### Reflections

From the summary above, it seems apparent that these new classes in applied technology were quite successful. As with all new projects, the program had its rough spots, but interviews with teachers and district administrators confirm the belief that the program is heading in the right direction.

The greatest challenge will be to continue to improve the scoring rubrics so that they clearly define the expectations for student performance. Except for the Work Habits Rubric, the assessment instruments lack specified criteria for the various performance levels. Data can be compiled in a multitude of ways. Only the teachers and students know which would be most useful to meeting their own goals. The examples provided are reasonable and useful only as they relate to the identified needs of the classrooms involved in developed them. As the rubrics are further developed and refined, however, they should be useful in other areas of the school or district. For example, the Presentation Skills Rubric could be developed and used by all staff who want students to develop those skills. Schoolwide use would also provide consistency across the curriculum. Some rubrics could be generic enough to be used at many different levels. Some specifics, of course, would apply only to certain content, but perhaps those skills would be judged with a different scoring device.

As the teachers and students work together and examine students' products and the presentations, discussions will lead to more specific and meaningful assessment procedures that exemplify standards for quality performance. That is definitely the intention and hope of those involved in this forward-looking approach to teaching and learning new skills in technology.

# VII-7

# Using Standards and Assessments to Promote Excellence and Equity Among Diverse Students

CLAUDE GOLDENBERG
CALIFORNIA STATE UNIVERSITY, LONG BEACH

BILL SAUNDERS
UNIVERSITY OF CALIFORNIA, LOS ANGELES

Over the past decade, educators and policymakers have increased the call for high academic standards and for assessments designed to gauge whether students achieve those standards. In response, national commissions, task forces, and policy groups have launched numerous efforts to develop ambitious goals for student learning and new ways of testing whether these goals are met.

Many, however, have expressed reservations. Who will set these standards and design the assessments? Will they be the same for all learners? Will they be equally valid for diverse groups of learners and at different levels of achievement? How will we assure equity for diverse students? (See, for example, Eisner 1993, Garcia 1993, and Viadero 1993.) Although we generally support many of the efforts to improve student learning through high standards and valid assessments, we share these concerns about equity.

No one seriously questions the need to promote excellence in U.S. schools. The real questions concern the meaning we attach to this idea: Excellence for whom? And how can we assure equal access for diverse groups of students? If we improve overall—or average—achievement without improving it at *all* levels (i.e., among lower- as well as higher-achieving students), then we will surely have failed to achieve one of the most important goals of the current reform movements.

---

This chapter was made possible by funds from the Spencer Foundation and from the Center for Research on Cultural Diversity and Second Language Learning, Office of Educational Research and Improvement, U.S. Department of Education. No endorsement from either source is implied nor should be inferred.

We have attempted to deal with these issues within the context of a project designed to improve academic achievement in an urban, largely Latino school in southern California. Our focus is not assessment per se but rather improving student learning and achievement. Although high standards and valid assessments are pertinent to this effort, they should be regarded as part of the concerted schoolwide effort to improve student learning. It is in the "context of use" that we think high standards and valid assessments will be most likely to produce results across the entire achievement spectrum.

### Goals and Assessments

Five years ago, collaborating with the school principal and several teachers, we began a multiyear project to improve academic achievement among all students at a low-income, low-achieving elementary school in southern California. We have been guided by a theoretical model that posits four essential elements for the change process (Sullivan 1990):

- **Goals** that are set and shared.
- **Indicators** that measure success.
- **Assistance** by capable professionals.
- **Leadership** that supports and pressures.

A fundamental premise was that these four elements together could be used to improve achievement at local school sites (Goldenberg and Sullivan 1994).

A key part of our work involved developing academic goals for all students (i.e., grade-level expectations developed with the entire faculty) and specific assessments to gauge whether students have accomplished these goals. We are interested in two aspects of student achievement: (1) Is student achievement improving overall? and (2) Is student achievement improving at all levels—that is, among lower- as well as higher-achieving students at the school?

### Assessments in Language Arts

We have seen broad-based improvements in student achievement, based on standardized tests of Spanish reading and state-mandated tests of English reading and writing (Goldenberg and Sullivan 1994). Our annual performance-based assessments also point to gains in the language arts. (We have so far concentrated on goals and assessments in language arts, with mathematics as our next targeted area.)

As agreed upon by the staff, a random sample of students (about 20 percent at each grade level) takes part in the language arts assessment at the end of each year. The results described here are based on assessments conducted in the first two years of our spring assessment administration. The assessment included 2nd through 5th graders who had to complete a series of tasks:

1. Reading Inventory. Students list books and stories they have selected and read on their own over the past year.
2. Self-Selected Story. Students choose a story they have read, summarize it, and then explain what they think is the most important part or theme.
3. Assigned Story. Students (grades 3–5 only) read a portion of a grade-level-appropriate story, summarize what they have read, and then write an ending for the story.
4. Assigned Article. Students (read a grade-level-appropriate informational text, synthesize what they learned, and then respond to an inference question specific to the text.
5. Dictation. Students take dictation for a grade-level-appropriate text.

In addition, we also score student papers (using the "self-selected story" task) for effective use of writing conventions such as spelling, punctuation, and appropriate language.

Staff from our project school and other schools in the area score the student products in an organized scoring session held on the first two days of summer vacation. They read papers without any identifying information about the student authors. Because teachers also score papers of students from different schools that have participated in our data-gathering efforts, the teachers have no way of knowing whether students are from their own or other schools.

### Improvements Across Achievement Levels

The spring assessments revealed gains in various areas across the grade levels. The strongest set of improvements emerged in grade 5, where statistically significant increases occurred in all areas for which we have valid comparisons for the two years—the number of books and stories reported on the Reading Inventory, summary and analysis of Self-Selected Story, Dictation, and use of written conventions.[1] (We do not report year-to-year comparisons for Assigned Story and for Assigned Article because baseline data for 5th graders on those two tasks were collected beginning in the second year.)

We then analyzed the 5th grade results on students' prior levels of achievement. Did the improvements we saw between 1992 and 1993 hold for both relatively high and low achievers? We defined "lower-achieving" students as those who scored below the school median on the fall administration of CTBS-U (California Test of Basic Skills), which in year

---

[1] All data reported here have been subjected to statistical tests with $p \leq .05$ as our level of significance.

1 was the 23rd percentile and in year 2 was the 29th; "higher-achieving" students scored above the median.

Both student groups improved significantly. Each group demonstrated significant gains from one year to the next on all areas of the assessment for which we have comparisons. Examples follow:

- Both higher- and lower-level achieving students listed more books. Higher-achieving students, who listed an average of 9 books in year 1, increased to an average of 16 in year 2. Lower-achieving students increased from 6 to 10.
- Both student groups revealed a higher level of understanding and wrote more clearly about a book or story they had read. The mean score for higher achievers improved from 2.67 to 3.44 (on a scale of 1-6); for lower achievers, means increased from 2.33 to 3.00.
- In both groups, increases were noted in the percentage of students who achieved the grade-level benchmark—a score point of 4. According to our scoring rubrics, a 4 means the student produced (1) a complete, understandable, and focused summary and (2) a clear articulation and explanation of a potential theme for the story. In year 1, only 11 percent of the higher achievers and none of the lower achievers received a 4. The following year, those percentages increased to 44 percent and 33 percent, respectively.

The picture was much the same for written conventions and dictation: Regardless of their prior levels of assessed achievement, 5th graders in year 2 performed at significantly higher levels than 5th graders the year before.

Although these improvements are encouraging, too many students are still not achieving at what we consider to be a desirable level of performance for their grade level, whether we judge performance in terms of standardized tests or our own locally developed assessments. Nonetheless, these data make us hopeful that we are headed in the right direction, in terms of both excellence *and* equity.

### Standards, Assessments, and Contexts of Use

We would argue that neither beneficial nor detrimental effects can be attributed to higher standards or more challenging assessments in and of themselves. Nor do we believe that standards and assessments inherently put low-achieving or diverse groups of students at further disadvantage. Rather, the critical factor is the context in which standards and assessments are used. If a context that promotes cohesive efforts to improve achievement for all students is absent, standards and assessments can *heighten* inequities and put students at risk, as many critics have warned.

The results we have reported emerged from a school context where the driving force is improving achievement for all students. Higher standards and more challenging assessments play a critical role, but they are not the only players. If they were, we would probably not see the achievement changes we are beginning to observe.

Several features of the school context help explain our results. First, staff, collaborating with researchers, developed both the grade-level goals and the assessments locally. Although other relevant work helped, the goals and assessments were not simply imported from other schools or policymaking agencies. Thus, the goals and assessments are locally meaningful and indicate levels of student performance that staff members generally accept.

Second, part of the impetus for developing goals and moving towards higher standards came from staff concerns about unacceptably low levels of achievement at the school (Goldenberg and Sullivan 1994). Our long-term efforts to improve student achievement were thus intrinsically meaningful and relevant to many teachers' day-to-day concerns.

Third, the goals and assessments continue to evolve in how they are understood, interpreted, and acted upon by the staff. Although the goals were finalized two years ago, and the assessments have been administered twice, debates and dialogues about them continue. Are the goals and assessments consistent with those that might be employed in other communities and with other populations of students? Are they leaving our lower-achieving students behind? How do we use the information these assessments provide to continue working towards higher levels of achievement?

Fourth, and as a result of the continued evolution of the goals and assessments and the questions they generate, an increasingly rich social, intellectual, and professional dynamic exists. With staff agreement, the principal has organized a number of opportunities for staff to discuss, make decisions, and learn new things. A standing group of grade-level representatives developed the goals. A subsequent group is developing classroom-appropriate assessments to be used on an ongoing basis and consistent with those used in the annual spring assessment. Each teacher participates in a weekly work group to share and discuss curricular and instructional strategies. (There are six such work groups: math problem solving, cooperative learning, instructional conversation, language arts, kindergarten teachers, and new teachers.)

Finally, staff members continue to feel both pressure and support to improve student achievement. The principal and the staff instructional specialist meet quarterly with teachers at each grade level to discuss student progress on the goals. The principal provides opportunities for the teachers to shape school efforts through such means as the goal-setting committee, assessment committee, grade-level meetings, and work groups. Pressure and support also come implicitly from the staff themselves, who hear how others are doing in their quarterly meetings, confer during work group sessions, conduct a random sample assessment at the end of each year, and score and then analyze the annual assessment results.

These features create a context where the emphasis is on improving student achievement at all levels. Standards and assessments help provide focus and ongoing information about how staff efforts are faring. They have become one of several tools that teachers use to help enhance educational opportunities for all students. In such a setting, equity will not take a back seat to excellence.

## REFERENCES

Eisner, E. (1993). "Why Standards May Not Improve Schools." *Educational Leadership* 50, 5: 22–23.

Garcia, E. (1993). "Linguistic Diversity and National Standards." *Focus on Diversity* 3, 1: 1–2.

Goldenberg, C., and J. Sullivan. (1994). *Making Change Happen in a Language-Minority School.* (EPR #13). Washington, D.C.: Center for Applied Linguistics.

Sullivan, J. (1990). "Elements for Successful Change." Unpublished manuscript, University of California, Los Angeles, Graduate School of Education.

Viadero, D. (June 2, 1993). "The Rhetoric and Reality of High Academic Standards." *Education Week*, pp. 16–17.

# VII-8

# Moving from Seat Time to Mastery: One District's System

KEVIN CASTNER, LORRAINE COSTELLA, AND STEPHEN K. HESS
FREDERICK COUNTY PUBLIC SCHOOLS

*The journey to become an outcome-based system is fraught with many lessons. Schools in Frederick County, Maryland, share a map of their process.*

Six years ago, Frederick County, Maryland, began a major restructuring effort, involving 30,000 students and 46 schools and covering all disciplines and grades. When we decided to become an outcomes-based school system, we could only guess at the impact our undertaking would have upon student learning—to say nothing about our curriculum or our philosophies on teaching and assessment!

The first generation (1988-92) of restructuring involved five components—some were planned and some forced themselves upon us: developing student outcomes, defining curriculum, developing new assessment tasks, planning for instruction, and providing for staff development throughout the process. The good news about restructuring is you can start tomorrow. The bad news is you are never finished.

## Our First Steps

The heart of what we call the "System for Effective Instruction" is a three-part essential curriculum pyramid model that creates objectives for each course, goals for each discipline, and—most important—five integrated learner behaviors that serve as developmental and exit outcomes (see Figure 1).

First, we had to define what Frederick County students should be able to do in each curricular area and how well they should be able to do it. The goal was to align our written, taught, and tested curriculum. At the top of our list was an emphasis on deep understanding and higher-level thinking skills.

Our five essential learner behaviors describe what each student will be able to do upon graduation in order to pursue lifelong learning and develop career and economic competence through:

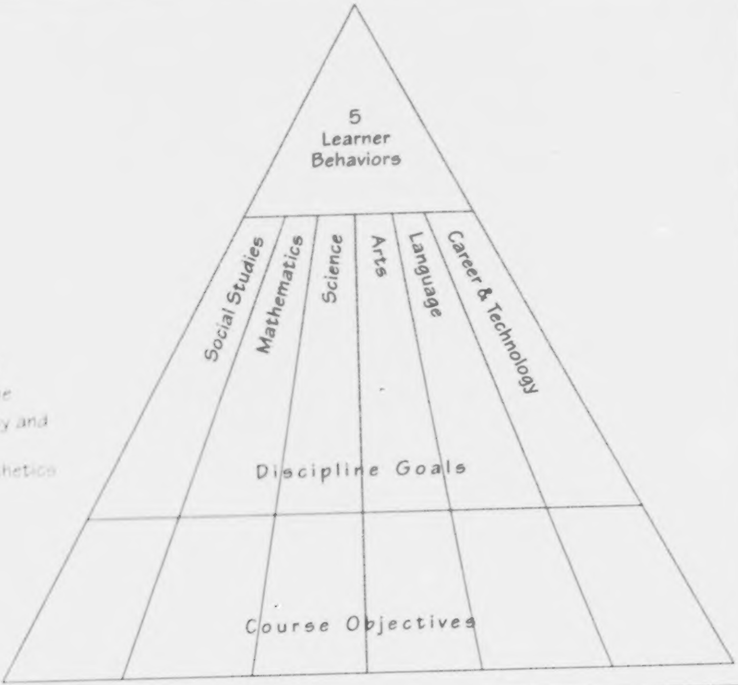
- effective communication,
- problem solving and critical thinking,
- social cooperation and self-discipline,
- responsible citizenship in community and environment, and
- lifelong learning.

---

Adapted from K. Castner, L. Costella, and S. Hess, "Moving from Seat Time to Mastery: One District's System, *Educational Leadership* 51, 1 (September 1993): 45-50.

Figure 1

ESSENTIAL CURRICULUM



**Essential Learner Behaviors**

1. Effective communication skills
2. Problem solving and critical thinking
3. Social cooperation and self-discipline
4. Responsible citizenship in community and environment
5. A lifestyle valuing wellness and aesthetics

## Demonstration of Learning

Next, we had to determine what we would accept as evidence of students' learning. To that end, two of our eight high schools participated in pilot projects to develop student portfolios and senior projects to indicate whether students were, indeed, achieving the essential curriculum.

For example, a senior project to show effective communication might be an oral presentation, a written communication, or a visual display. A student could demonstrate problem-solving and critical thinking skills by defending solutions to real-life problems and situations. The ability to assess self, others, and ideas demonstrates critical thinking skills.

In the pilot, each student, assisted by a teacher, was encouraged to develop original projects if he or she wished. A project with a career and technology focus might be the following:

> You are a carpenter assigned the responsibility for working with a customer to determine the best design, as well as estimated costs, for building a garage on a property with an existing residence. You must:
>
> - interview the customer to determine storage needs, cost allowances, and preferred exterior finishes;
> - develop a proposal including recommended design, exterior finish, and a basic sketch;
> - write a cover letter to your customer summarizing your proposal and explaining why your firm should be hired to do the work.

In each curricular area, essential learner behaviors are supported by essential discipline goals, which, in turn, are supported by essential course objectives. For instance, a task that requires 7th graders to plan a field trip to a museum in Washington, D.C., could meet two essential course objectives: (1) collect, organize, represent, and interpret data and (2) make estimates appropriate to given situations.

These 7th grade objectives support our K-12 mathematics discipline goal: to develop mathematical skills and reasoning abilities needed for problem solving. In addition, the lesson helps students gain skills in effective communication, social cooperation, and citizenship.

Each level and grade of schooling, beginning in kindergarten, uses the foundation of individual

BEST COPY AVAILABLE

courses and disciplines to build toward mastery of the learner behaviors at the top of the pyramid. Integrated essential learner behaviors are important for success in life. No one in the real world stops and says, "Now I am going to do Algebra I, AP Chemistry, or World History." They just switch gears to accomplish a task.

## New and Improved Assessments

What about testing? The ability to fill in the correct answer bubble is not enough for our new curriculum. True mastery requires the ability to apply what has been learned to real-world situations.

For example, an essential course objective in 5th grade language arts is for a student to read or listen to an expository selection in order to process information at the literal, interpretive, and critical levels. Our emphasis was on thoughtful mastery of important tasks, rather than thoughtless knowledge of isolated facts and skills.

Traditionally, because of ease of measurement, our assessment tasks had tended to limit instructional focus to lower-level skills and knowledge. The clear intent of the Frederick County essential curriculum was to reverse this tradition. We determined what was important for students to know and be able to do, and then we decided how to measure it. This system requires the performance of exemplary tasks that reflect the actual performances expected of students.

If we want students to "think with words and think with numbers," we must emphasize the evaluation of student ability to understand and problem solve. The isolated reading passages and simple word problems found on standardized tests did not meet this need. Routine assessment of prerequisite skills has a place in our criterion-referenced evaluation system, but the relative balance is weighted in favor of critical thinking, problem solving, and communication.

As we began developing activities, we emphasized formative assessment to give teachers a way to monitor student progress on the essential curriculum. We view formative assessments not as teaching to the test, but testing what we teach.

Although initial training, feedback, and review were provided by experts like Grant Wiggins, Frederick County teachers actively developed the assessment tasks. In February 1988, when we invited teachers to apply for summer curriculum/assessment writing workshops, their response was outstanding. As one teacher said, "I don't want someone else telling me what I'll be teaching for the next 10 years." In designating teachers for the workshops, curriculum specialists sought a balanced representation: experienced and new teachers, males and females, cultural diversity, and a range of instructional levels.

In the spring, acceptance letters went out, emphasizing the importance of the workshops and the work done in them to our school system. Planning for the summer workshops included extensive research and visits and phone calls to school systems already involved in performance-based assessment. Before we began work, we took advantage of several opportunities to plan and confer with consultants.

## The Summer Workshops

Two weeks of workshops began with much fanfare that summer. More than 200 teachers from grades K-12 attended the initial summer workshop. The superintendent and the president of the board of education were there to set the tone and talk about the importance of assessments in our restructured school system. Participants with young children found child care available on the premises. A media specialist provided research materials, and on-site typists at a word processing center turned out finished products.

During the first day or two, several large-group meetings set the framework. Grant Wiggins launched the workshop series by providing an overview of authentic assessment. Then teachers and curriculum specialists worked to translate this broad perspective into real assessments for Frederick County.

We learned a great deal in our earliest workshop. As teachers created assessment tasks, their efforts touched other curriculum areas and grade levels. What developed was something unexpected and exciting—a common educational focus! However rosy a picture we paint, we admit that writing these first assessment tasks was a struggle. Teachers worked diligently to write the assessments, only to find that they were uninteresting to students, or that they did not test what we wanted them to test.

Finally, at the end of the first two-week workshop, the assessments began to jell. And when the workshops ended, each curricular area had several assessments to pilot.

As teachers were writing assessment tasks, they were also developing scoring systems for student performance at various levels. After debating the merits of different frameworks, we finally decided to use four-point rubrics for continuity across schools, grades, and disciplines. While specific rubrics are written for narrative writing, problem solving, and reading comprehension, for example, all use the same framework:

4 = exemplary (*fully* meets criteria)
3 = proficient (*adequately* meets criteria)
————Mastery Line————
2 = approaching proficiency (*sometimes* meets criteria)
1 = evidence of attempt (*seldom* meets criteria)

The four points are not equivalent to letter grades. They are a method of assessing the degree of proficiency exhibited in a student's response to an assessment task. The goal is to have all students progress to the point of mastery, with the teacher functioning in the role of a coach.

Our goal for the initial year was to pilot assessments in language arts and math in grades K-8. Eventually we hope to develop essential curriculum and assessment tasks in every area. To date, we have accomplished this in language arts, math, science, social studies, art, music, physical education, and business education.

## Selection of "Anchor Papers"

Once the teachers had completed their work on assessments during the workshops, our curriculum specialists spent the remainder of the summer finalizing the tasks and planning teacher training. During a day of inservice before schools opened, the curriculum specialists met with teachers in their curricular areas and presented overviews of performance-based assessments, discussed the importance of formative assessments, and illustrated tasks that had been written during the summer.

Teachers in the pilot schools worked closely with the curriculum specialists to collect examples of student work. These examples, called *anchor papers*, provided consistency in evaluating student work and gave us samples of mastery, as well as every other point on the rubric.

The selection of anchor papers took place at a number of painful meetings during the first four months of school. Selection was the result of serious debate, which made us look at what the standards for mastery were. For example, teachers asked, "Is this what we want the writings of a 3rd grader to look like?" or "What needs to be in a response that shows mastery?" It was very difficult to look only at responses and not focus on the effort of the student.

By looking at the anchor papers, we found that teachers could learn how to adjust instruction. After piloting the assessments and agreeing on the anchors that best illustrated the standards, we faced the task of inservicing all the other teachers! The only way we could accomplish this awesome task was to use a "trainer of trainers" model. In elementary schools, the language arts curriculum specialist met with the reading specialists and developed a training module on assessment. Each reading specialist took the module and used it as the basis for staff development in his or her school.

Variations on this procedure were used in every curricular area. Sometimes curriculum specialists worked with department chairs, team leaders, or assistant principals. These people, in turn, worked with the teachers. Of course, curriculum specialists, our testing supervisor, and staff development facilitators were available for assistance.

Frederick County has used the following model of curriculum and assessment development for the last five years:

1. write the curriculum and assessment;
2. pilot the curriculum and assessment in selected schools;
3. refine and adjust the curriculum and assessment;
4. provide inservice for all teachers; and
5. adjust, refine, and rewrite.

The curriculum can be changed and the assessments can be changed, but to make an impact on instruction and achieve higher-level student behaviors, the change must take place in instruction.

## Improvement as a System

In our school system, yearly progress reports are published so that our public can view the results of these assessments. Our community is informed each August of the percent of students who have mastered the essential curriculum, with the results disaggregated by gender, ethnicity, and socioeconomic status. A copy of a page from the most recent progress report is provided in Figure 2. Individual school improvement teams in each county school meet to review progress and plan for further improvement.

Implementation of the Maryland School Performance Assessment Program (MSPAP) followed the development of our local CRES program. The two programs are similar in intent and design, reflecting a future-directed standard for student achievement and emphasizing the use of state-of-the-art performance assessments. A similar public reporting system is used (see Figure 3).

In marked contrast to traditional conflicts between state and local testing programs, the parallels between the two programs enable us to use the state results as an external validation of our locally developed assessments. Teachers and administrators must be confident that student success with our local essential curriculum and CRES assessments will lead to success on statewide testing programs.

Through intensive statistical review of the two programs, MSPAP results were compared with local CRES mastery percentages. Given the fact that local assessments included additional and prerequisite objectives, the percentages were not expected to exactly agree. Indeed, the state goals for schools is to have 70% (or more) of students scoring in the "satisfactory" range by 1996, while the goal for our local CRES assessments has been to eventually have 90% (or more) of students mastering the assessments. Considering the differences, it appeared reasonable to expect similar differences from these goals if the two

FIGURE 2

## FREDERICK COUNTY PUBLIC SCHOOLS

## *GOAL A*

***(System Goal —90% of students achieve for each indicator)***

CRES SUMMATIVE ASSESSMENTS (LANGUAGE ARTS) % AT/ABOVE 80% MASTERY (GRADES 1-3)

| | 91-92 | 92-93 | 93-94 |
|---|---|---|---|
| All | 81% | 80% | |
| Male | 77% | 76% | |
| Female | 84% | 83% | |
| White | 83% | 81% | |
| African American | 62% | 62% | |
| M/H SES | 83% | 82% | |
| Low SES | 52% | 55% | |

CRES SUMMATIVE ASSESSMENTS (LANGUAGE ARTS) % AT/ABOVE 80% MASTERY (GRADES 4-5)

| | 91-92 | 92-93 | 93-94 |
|---|---|---|---|
| All | 83% | 83% | |
| Male | 79% | 79% | |
| Female | 88% | 87% | |
| White | 85% | 85% | |
| African American | 67% | 63% | |
| M/H SES | 85% | 86% | |
| Low SES | 57% | 53% | |

CRES SUMMATIVE ASSESSMENTS (LANGUAGE ARTS) %AT/ABOVE 80% MASTERY (GRADES 6-8)

| | 91-92 | 92-93 | 93-94 |
|---|---|---|---|
| All | 59% | 59% | |
| Male | 52% | 51% | |
| Female | 66% | 68% | |
| White | 60% | 61% | |
| African American | 36% | 33% | |
| M/H SES | 62% | 62% | |
| Low SES | 30% | 31% | |

CRES SUMMATIVE ASSESSMENTS (LANGUAGE ARTS) % AT/ABOVE 80% MASTERY (GRADES 9-12)

| | 91-92 | 92-93 | 93-94 |
|---|---|---|---|
| All | 58% | 57% | |
| Male | 49% | 49% | |
| Female | 68% | 66% | |
| White | 59% | 59% | |
| African American | 37% | 29% | |
| M/H SES | 60% | 60% | |
| Low SES | 40% | 37% | |

CRES SUMMATIVE ASSESSMENTS (MATHEMATICS) % AT/ABOVE 80% MASTERY (GRADES 1-3)

| | 91-92 | 92-93 | 93-94 |
|---|---|---|---|
| All | 89% | 89% | |
| Male | 88% | 88% | |
| Female | 90% | 90% | |
| White | 91% | 90% | |
| African American | 74% | 76% | |
| M/H SES | 91% | 91% | |
| Low SES | 72% | 73% | |

CRES SUMMATIVE ASSESSMENTS (MATHEMATICS) % AT/ABOVE 80% MASTERY (GRADES 4-5)

| | 91-92 | 92-93 | 93-94 |
|---|---|---|---|
| All | 85% | 84% | |
| Male | 83% | 82% | |
| Female | 87% | 85% | |
| White | 86% | 85% | |
| African American | 69% | 71% | |
| M/H SES | 87% | 85% | |
| Low SES | 66% | 65% | |

CRES SUMMATIVE ASSESSMENTS (MATHEMATICS) % AT/ABOVE 80% MASTERY (GRADES 6-8)

| | 91-92 | 92-93 | 93-94 |
|---|---|---|---|
| All | 65% | 72% | |
| Male | 63% | 69% | |
| Female | 67% | 75% | |
| White | 67% | 74% | |
| African American | 34% | 49% | |
| M/H SES | 68% | 75% | |
| Low SES | 36% | 43% | |

CRES SUMMATIVE ASSESSMENTS (MATHEMATICS) % AT/ABOVE 80% MASTERY (GRADES 9-12)

| | 91-92 | 92-93 | 93-94 |
|---|---|---|---|
| All | 45% | 39% | |
| Male | 45% | 40% | |
| Female | 46% | 39% | |
| White | 47% | 41% | |
| African American | 25% | 17% | |
| M/H SES | 47% | 41% | |
| Low SES | 30% | 20% | |

CRES SUMMATIVE ASSESSMENT % ON-GRADE LEVEL, END OF GRADE 3 (LANGUAGE ARTS AND MATHEMATICS)

| | Language Arts | | Mathematics | |
|---|---|---|---|---|
| | 91-92 | 92-93 | 91-92 | 92-93 |
| All | 83% | 81% | 85% | 85% |
| Male | 79% | 78% | 84% | 84% |
| Female | 86% | 84% | 86% | 86% |
| White | 85% | 83% | 87% | 86% |
| African American | 64% | 61% | 69% | 67% |
| M/H SES | 85% | 84% | 87% | 87% |
| Low SES | 57% | 57% | 64% | 66% |

BEST COPY AVAILABLE

FIGURE 3

# FREDERICK COUNTY PUBLIC SCHOOLS

Frederick County, with a population of 164,503, is Maryland's largest county in land mass. The Board of Education is housed in Frederick City, the County seat. The forty-seven public schools include twenty-seven elementary schools, ten middle schools, seven high schools, one special education center, one alternative education middle/high school, and one career and technology center. Outdoor school facilities are also part of the school system.

## ASSESSED STUDENT KNOWLEDGE

| MD Functional Tests Grade 9 Status | Standard % | | 1993* | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | EX | SAT | No. Taking | No. Absent | No. Exempt | % Passing | EX | SAT | Not Met |
| Reading | 97 | 96 | 2,049 | 32 | 0 | 98.6 | ✓ | | |
| Mathematics | 90 | 80 | 2,063 | * | 0 | 92.8 | ✓ | | |
| Writing | 96 | 90 | 1,998 | 83 | 0 | 97.2 | ✓ | | |
| Citizenship | 92 | 85 | 1,901 | 25 | 0 | 94.2 | ✓ | | |

| MD Functional Tests Grade 11 Status | Standard % | | 1991 G | 1992 G | 1993* | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | EX | SAT | | | No. Refused | No. Exempt | % Passing | EX | SAT | Not Met |
| Reading | 99 | 97 | 99.9 | 99.6 | 0 | 0 | 99.9 | ✓ | | |
| Mathematics | 99 | 97 | 99.3 | 99.0 | 0 | 0 | 98.4 | | ✓ | |
| Writing | 99 | 97 | 99.5 | 99.2 | 0 | 0 | 99.3 | ✓ | | |
| Citizenship | 99 | 97 | 98.1 | 98.4 | 0 | 0 | 98.4 | | ✓ | |
| Passed All Tests | 96 | 90 | 97.3 | 97.4 | 0 | 0 | 97.2 | ✓ | | |

| MSPAP 1992 RESULTS | Proposed Standard | | No. Tested | No. Absent/ Excused | No. Not Reported | Percent at | | EX | SAT | Not Met |
|---|---|---|---|---|---|---|---|---|---|---|
| | EX | SAT | | | | EX | SAT | | | |
| **GRADE 3** | | | | | | | | | | |
| Reading | 25 | 70 | 2,050 | 97 | 110 | 5.3 | 44.4 | | | ✓ |
| Mathematics | 25 | 70 | 2,094 | 110 | 53 | 2.9 | 40.7 | | | ✓ |
| Social Studies | 25 | 70 | 2,070 | 134 | 53 | 5.0 | 46.5 | | | ✓ |
| Science | 25 | 70 | 2,071 | 133 | 53 | 3.9 | 44.7 | | | ✓ |
| **GRADE 5** | | | | | | | | | | |
| Reading | 25 | 70 | 1,974 | 125 | 105 | 8.2 | 40.2 | | | ✓ |
| Mathematics | 25 | 70 | 2,043 | 115 | 46 | 11.2 | 56.9 | | | ✓ |
| Social Studies | 25 | 70 | 2,014 | 144 | 46 | 7.7 | 45.9 | | | ✓ |
| Science | 25 | 70 | 2,008 | 150 | 46 | 4.6 | 48.0 | | | ✓ |
| **GRADE 8** | | | | | | | | | | |
| Reading | 25 | 70 | 1,683 | 151 | 83 | 4.6 | 36.3 | | | ✓ |
| Mathematics | 25 | 70 | 1,730 | 145 | 42 | 8.1 | 48.8 | | | ✓ |
| Social Studies | 25 | 70 | 1,719 | 156 | 42 | 6.6 | 44.2 | | | ✓ |
| Science | 25 | 70 | 1,694 | 181 | 42 | 4.5 | 41.3 | | | ✓ |

*Fewer Than 20 Students **Indicates Baseline Year Data

***Includes Exemptions for Special Education, ESOL, and Second Semester Transfers

KEY: EX = Excellent, SAT = Satisfactory % = Percent MSPAP = Maryland School Performance Assessment Program

H.S. = High School

## STUDENT PARTICIPATION

| Attendance Rate Yearly | Standard % | | 1990** % | 1992 % | 1993 | | | |
|---|---|---|---|---|---|---|---|---|
| | EX | SAT | | | % | EX | SAT | Not Met |
| Grades 1-6 | 96 | 94 | 95.0 | 95.3 | 95.3 | | ✓ | |
| Grades 7-12 | 96 | 94 | 92.0 | 92.7 | 92.7 | | | ✓ |

| Students Absent | 1992 Percent | 1993 Percent |
|---|---|---|
| Fewer than 5 days | 30.8 | 32.1 |
| More than 20 days | 11.0 | 10.7 |

| Dropout Rate Yearly | Standard % | | 1990** % | 1992 % | 1993 | | | |
|---|---|---|---|---|---|---|---|---|
| | EX | SAT | | | % | EX | SAT | Not Met |
| Grades 9-12 | 1.25 | 3.0 | 2.9 | 1.8 | 1.79 | | ✓ | |

## STUDENT ATTAINMENT

| Promotion Rate | Standard % | | 1990** % Promoted | 1992 % Promoted | 1993 | | | |
|---|---|---|---|---|---|---|---|---|
| | EX | SAT | | | % Promoted | EX | SAT | Not Met |
| Grades 1-6 | 98 | 96 | 99.6 | 99.8 | 99.5 | ✓ | | |

| High School Program Completion | 1991 Percent Completed | 1992** Percent Completed | 1993 | |
|---|---|---|---|---|
| | | | Number Completed | Percent Completed |
| University of Maryland System Requirements | 49.6 | 48.8 | 888 | 50.9 |
| Approved Occupational Program Requirements | 12.4 | 16.7 | 288 | 16.5 |
| BOTH-University & Occupational Requirements | 3.5 | 2.2 | 41 | 2.3 |

## POST-SECONDARY DECISIONS

| Grade 12 Documented Decisions to: | 1992* | 1993 | |
|---|---|---|---|
| | Percent | Number | Percent |
| Attend a four year college | 36.6 | 552 | 36.2 |
| Attend a two year college | 21.8 | 309 | 20.2 |
| Attend a specialized school or pursue specialized training | 1.9 | 43 | 2.8 |
| Enter employment (related to H.S. program) | 8.2 | 130 | 8.5 |
| Enter employment (unrelated to H.S. program) | 9.7 | 160 | 10.5 |
| Enter the military | 3.6 | 82 | 5.4 |
| Enter full-time employment and school | 2.7 | 39 | 2.6 |
| Enter part-time employment and/or school | 10.0 | 167 | 10.9 |
| Other and no response | 5.5 | 44 | 2.9 |

BEST COPY AVAILABLE

programs were complementary. This "70-90" concept was used to identify content areas or grades where the *disparity* from the goals substantially varied, as illustrated in figure 4.

As is readily apparent, while the actual percentages vary, the degree of difference from the state and local goals is similar for reading and math. A fifteen percent gain in mathematics would meet the goals for both respective programs, as would a 33% gain in reading. The pattern in social studies is quite different, in that the 13 point gain required to meet the CRES goal would be insufficient to meet the goal for the state assessment. This pattern was used to revise the CRES social studies assessments so that the standards for the two programs were no longer in conflict.

A final "check" on the data was to identify patterns of *relative* performance on the two assessment programs for individual schools, in the vast majority of cases, schools with relatively high CRES percentages had high MSPAP percentages, and conversely, schools with relatively low CRES percentages had low MSPAP scores. In a few cases, schools had MSPAP scores out of range with CRES percentages, suggesting the possibility of "effort scoring" or the need for additional scoring training.

The yearly randomized checks of scoring consistency were adjusted to oversample the few schools so identified, so that teachers and administrators could see the results of increased training.

## What We've Learned

After six years of hard work, we realize the key role staff development plays in the restructuring process. Teachers need to learn how to assess students' knowledge and also how to teach the higher-level behaviors required by the performance tasks. The change from multiple-choice testing to teaching that emphasizes problem solving and higher-order thinking cannot be assumed. It must be planned.

Another lesson is that the development and scoring of assessments is labor-intensive. Time is needed to write assessment tasks and to train teachers how to give and score them. We have found it very profitable to combine resources with other districts in order to develop assessments and share training of staff. During the summer of 1991, 16 Maryland school districts worked together to develop assessments, rubrics, and anchors for language arts, mathematics, science, and social studies in grades K-8. The result was three volumes of assessments, which would have taken each of us years to develop! The assessments are now part of a computerized database, which is accessible to the staff upon request. The Maryland Assessment Consortium, an ongoing entity, represents all 24 Maryland school districts.

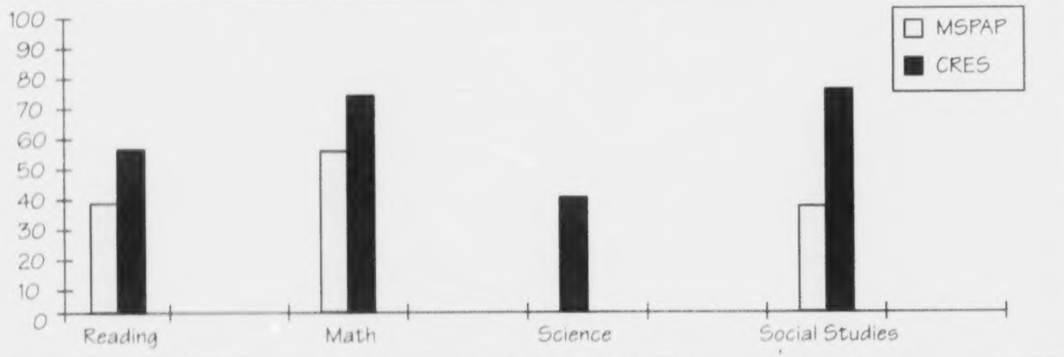
While much attention has been given to the written curriculum and assessments, if assessment is to increase learning, it must be directly integrated into daily teaching. Only when teachers learn to analyze the assessments will they be able to coach students in the higher-level behaviors.

## What Lies Ahead

The second generation (1992-97) of our restructuring process is already under way. A videotape and brochure have been provided to all schools to assist parents and the community in understanding "Teaching and Testing for the 21st Century."

The original design of the System for Effective Instruction emphasized building the foundation of the essential curriculum pyramid. The second generation will focus more directly on the original reason for moving from seat time to mastery: ensuring that all students reach high standards of understanding and application. Developmental and exit portfolios will eventually provide the evidence of integrated mastery required for success as productive adults and lifelong learners.

FIGURE 4
PATTERNS OF PERFORMANCE BY CONTENT AREA



# VII-9

# Using the Results of Performance Assessment to Support School Reform and Improve Student Achievement

LOUISE F. WAYNANT
PRINCE GEORGE'S COUNTY PUBLIC SCHOOLS, MARYLAND

A visitor to any of the 150 elementary and middle schools in Maryland's Prince George's County Public Schools during the month of May would observe students intently involved in experiments, investigations, and small-group discussions. Students would be charting information, writing extended responses to a variety of questions, comparing and contrasting information, and participating in similar exciting, "hands-on" activities. At first, it may appear that students are involved in a special class project. Actually, they are participating in a form of high-stakes, state-mandated accountability testing known as the Maryland School Performance Assessment Program. Results from these assessments, given annually to all Maryland students in grades 3, 5, and 8, are used to determine if schools and school systems across the state have met challenging new state accountability standards. Schools not meeting standards or showing progress toward reaching the standards may be recommended for state intervention termed "reconstitution."

These high-stakes performance assessments have dramatically affected instruction throughout Maryland's schools. In the Prince George's County Public Schools, performance assessments have prompted significant changes in curriculum, instruction, assessment, instructional materials, and staff development. These changes have helped the school system attain high levels of participation and achievement for all students.

## A Mission of School Reform

The Prince George's County Public Schools is the 15th largest school system in the United States, serving more than 116,000 students from diverse backgrounds. The student population is 70.3 percent

African American, 20.2 percent white, 5 percent Hispanic, 4.1 percent Asian, and .4 percent Native American. Adjacent to Washington, D.C., the school system experiences many of the challenges confronting urban school districts in the 1990s.

The Prince George's County Public Schools began its school reform initiative during the 1985–86 school year by implementing the Effective Schools Process for School Improvement. Subsequently, the school system adopted a mission statement that targets the attainment of high levels of achievement of both basic and higher-order skills by all students. Six comprehensive school system goals guide the school reform effort:

- Student achievement will meet or exceed state and national standards in all curricular areas, and this achievement will reflect the application of higher order thinking skills.
- Any gaps in achievement among groups of students identified by ethnicity, gender or poverty will close, while achievement for all students will increase.
- School system curricula and learning climate will fully reflect and support a multicultural perspective.
- School climate and facilities will be safe, orderly and supportive of teaching and learning.
- Parents and the community will be meaningfully and consistently involved in the educational process and the decisions which shape it.
- The school system will be fully accountable for all of its programs, practices, and results.

The school system has developed a comprehensive strategic plan that guides and informs the school renewal process, identifying specific objectives, strategies, and time lines chosen to fulfill its mission and achieve its goals. Among these strategies is the use of performance assessment for formative and summative evaluation purposes.

## Performance Assessment and School Reform

Performance assessment is an especially helpful tool in creating effective school reform, because it calls on students to *demonstrate* what they know and are able to do, requiring them to apply skills and concepts from a variety of disciplines to solve problems and create responses. All students must develop and apply higher-order thinking skills as they complete performance tasks as part of both the formative and summative assessment process. Perhaps most important, *all* students are expected to demonstrate and apply higher-level skills. It is this emphasis on high expectations for *all* students, and the shaping of curriculum, instruction, and assessment for *all* students toward that end, that are such powerful factors in promoting school reform.

## Types of Performance Assessment Used

Performance assessment in the Prince George's County Public Schools includes both formative and summative instruments developed by the Maryland State Department of Education and the Prince George's County Public Schools. These instruments are described below.

### The Maryland School Performance Assessment Program

The Maryland School Performance Assessment Program (MSPAP) is the assessment component of the Maryland School Performance Program implemented by the Maryland State Department of Education (MSDE) primarily for the purpose of school improvement. These criterion-referenced assessments consist solely of open-ended, constructed-response items. The tests include integrated performance tasks designed to measure students' ability to apply higher-order skills in reading, writing/language usage, mathematics, social studies, and science. The tests are administered annually in May. Students work in cooperative groups to collect data and then individually respond to highly challenging tasks, some of which continue over a period of several days.

Completing the assessments at each grade level requires nine and a half hours of testing over five days. Student response booklets are sent to MSDE, and during the summer months, specially trained teams of Maryland teachers score the responses according to previously defined rules and rubrics. Student performance in each of the content areas is assigned one of five proficiency levels, Level 1 being the highest. Results are reported as the percent of students whose scores fall within each proficiency level and as the percent of students achieving at state-defined "excellent" and "satisfactory" standards. Results are disaggregated by race and gender. (See the Appendix.)

### Locally Developed Criterion-Referenced Tests

For almost ten years, the Prince George's County Public Schools have developed and used criterion-referenced tests (CRTs) in reading, language arts, mathematics, social studies, and science for both formative and summative evaluation purposes. Initially, these tests were multiple-choice, with an emphasis on basic skills. With the advent of performance assessment as part of the state's accountability structure, the locally developed CRTs in reading/writing/language usage and mathematics have been revised to include a performance assessment component. Student responses are scored holistically by teams of system-trained teachers using previously developed scoring rules and rubrics. Results are reported to schools as individual student scores and as the percent of students achieving at defined point-score levels. Results are disaggregated by race and sex.

### Formative Assessment

In addition to locally developed and state performance assessment tasks, schools have a variety of sample performance tasks that can be modified and incorporated into daily classroom instruction. For example, the Maryland Assessment Consortium annually brings together select Maryland teachers for two weeks each summer to create and refine a variety of performance assessments for elementary, middle, and high school grades. The resulting sets of performance assessment tasks are available to all schools in the state. Teachers are encouraged to use these assessments as an ongoing, integral part of their instructional program, not solely as preparation for the state and local criterion-referenced tests described above.

### Teacher-Developed Classroom Assessments, Portfolios, and Projects

Teachers are encouraged to develop and use their own performance tasks in conjunction with their classroom instructional units and lesson plans. Other types of performance-based or authentic assessments, such as portfolios and projects, are encouraged also, and have been initiated by many schools. For example, some elementary schools use portfolio assessment to monitor student progress in reading/language arts and science. Items included in the portfolios include projects, reports, book summaries, taped reading performances, and writing samples. A major feature of portfolio assessment is an emphasis on students' metacognitive processing. Portfolios capture not only the products or projects students create, but their decision making and reflection on their work as well.

## Statewide Plan for Using Performance Assessment to Improve Student Achievement

To ensure that the use of performance assessments is fully embedded in the instructional program of the Prince George's County Public Schools and aligned with curriculum and instruction throughout the system, a *Statewide Plan for Using Performance Assessment to Improve Student Achievement* has been developed. Components of the plan include:

- Establishing the relationship of performance assessment to curriculum.
- The use of performance assessment for formative and summative purposes.
- Defining information and skills related to performance assessment that all teachers should have.
- Integrating performance assessment tasks into the curriculum of the Prince George's County Public Schools.
- Identifying resources and references related to performance assessment.
- Providing staff development opportunities related to performance assessment.
- Establishing the relationship of performance assessment to other assessments used in the school system.
- Correlating the relationship of performance tasks and related assessments to the Dimensions of Learning framework (see Marzano 1992).
- Identifying the instructional implications of performance assessment.

The plan provides school-based staff with the information they need to implement performance assessment. Educators also use it to monitor how effectively systemwide performance assessment is used in instructional program design, implementation, evaluation, and modification.

## The Instructional Effects of Performance Assessment

Performance assessment has profoundly affected the instructional program of the Prince George's County Public Schools. Systemwide results from state and local assessments highlighted the need for significant changes in both curriculum and instruction. Among these changes are those described below.

### Curriculum

1. **Learning outcomes** in major curricular areas have been revised to be consistent with outcomes measured on state assessments. For example, learning outcomes in mathematics are now aligned with the Standards of the National Council of Teachers of Mathematics as well as state mathematics outcomes. Learning outcomes for reading reflect students' ability to construct meaning, to read to be informed, to read to perform a task, and to read for personal enjoyment. Learning outcomes for social studies, science, and English/Language Arts also have been revised.
2. **Curriculum** at all levels is being modified to emphasize cross-disciplinary connections, with an emphasis on students' ability to apply knowledge and skills in problem-solving situations.
3. **Performance tasks** that match units and themes embedded in curricular materials are being developed to assist teachers in using formative assessment for classroom instruction.

### Instruction

1. Because performance tasks involve the application of higher-order thinking skills, teachers throughout the school system are encouraged to include **higher-order questions** in every lesson. Student responses are shared with peers and with the teacher through use of strategies such as "Think, Pair, Share," "Jigsaw," and "K-W-L."

2. **Activity-oriented lessons** that include opportunities for students to collect, organize, discuss, display, and interpret data are emphasized in a variety of content areas such as science, mathematics, and social studies, as well as through interdisciplinary instructional units. Students are encouraged to make predictions based on data collected and to summarize their work in graphic and written form.

3. Students are encouraged to daily **communicate through writing** in all content areas. The emphasis is on a process approach that includes prewriting, drafting, revising, and editing. Because responses to performance tasks in state and local assessments rely heavily on writing, students are taught strategies for revising and editing their writing. Teachers demonstrate or "model" these strategies during direct instruction and often provide guided practice. Students keep **journals** and **logs** in which they reflect on their learning.

4. Performance assessment results pointed to the need to have students **work more independently** and be able to **follow multi-step directions** without teacher intervention. This type of independence can be fostered gradually as students learn to **monitor their own thinking** and review their work against a set of standards.

5. Students have many opportunities to **work in cooperative groups** to investigate a problem, discuss a topic or issue, or gather and record data. Students also reflect, write, or develop a product independently, based on group work.

6. In working with mathematics problems and tasks, students extend their thinking and expand their work by **describing how they arrived at their answers**, the steps they followed, and the reasons they took those steps.

7. Students have many opportunities to develop skills in **summarizing and paraphrasing information** through direct instruction and guided practice.

8. Teachers use performance tasks for **formative assessment** to monitor student progress and adjust instruction.

9. Staff development ensures that teachers know how to use **holistic scoring procedures** as part of classroom instruction and assessment. Teachers help students use holistic scoring in assessing their work and in collaborating with peers to assess group projects and tasks. Students are involved in the development of evaluation criteria as an ongoing aspect of instruction. In this way, students develop self-evaluation skills as they review and reflect on their work.

## Issues Related to Performance Assessment

Although performance assessment is an important component of the school system's strategic plan, several related issues must be considered so they will not diminish the potential of performance assessment to improve student achievement:

- Aligning curriculum, instruction, and assessment to ensure that what is assessed is what is taught, and to ensure that daily instruction includes hands-on problem-solving tasks, higher-order questions, and opportunities to generate alternative solutions to real-life problems.
- Providing sufficient staff development so that teachers can develop, use, and interpret the results of performance assessment.
- Providing time for teachers to develop, score, and analyze performance tasks on an ongoing basis in their classrooms.
- Developing fair and useful scoring rubrics and rules for summative assessments
- Providing funds to support the scoring of performance assessments used for summative purposes.
- Providing the results of summative performance assessments to teachers in a timely manner.
- Linking performance assessment results with instructional interventions.
- Developing cost-effective procedures for creating, scoring, and reporting performance assessment results in a timely manner.
- Collecting evidence to validate the inferences drawn from students' responses to performance assessment.

## Summary

The good news about incorporating performance assessment in a strategic plan for school reform is that performance assessment can influence daily instructional practice in ways that create challenging, exciting, and meaningful learning environments for students. Performance assessment can also foster the application of higher-order skills in a variety of problem-solving situations. Ultimately, performance assessment can contribute significantly to the mission of schooling: attaining high levels of success for all students.

### REFERENCE

Marzano, R.J. (1992). *A Different Kind of Classroom: Teaching with Dimensions of Learning*. Alexandria, Va.: ASCD.

# APPENDIX
## Examples of Formats for Reporting the Results of Performance Assessment

### MSPAP Standards Reports (Appendix A)

These reports provide information relevant to the school performance mission of the Maryland School Performance Program (MSPP). They report percentages of students at satisfactory and excellent levels of performance and indicate whether the standards for satisfactory and excellent school performance have been met.

### MSPAP Proficiency Level and Participation Reports (Appendixes B, C, D)

These reports provide information relevant to the school improvement mission of MSPP. Proficiency-level reports for all students in a school, school system, and the state indicate the number and percentages of test takers at each of the five MSPAP proficiency levels. They also report the numbers and percentages of students who completed assessment activities in each MSPAP content area and therefore received a scale score. Finally, they report the numbers and percentages of students who were absent, excused, or exempted from the 1993 MSPAP test administration.

### MSPAP Outcome Score Reports (Appendix E)

Outcome Score Reports contain the average outcome score, or percentage of mastery of an outcome, for a school, school system, or the state. They also include percentages of students in four outcome score ranges: 0–25, 26–50, 51–75, and 76–100. This information is intended to provide a general idea of the percentage of students who have displayed little or no mastery of the knowledge, skills, and processes required in an outcome (i.e., those in the outcome score range 0–25) and the percentage who have displayed near complete mastery of the outcome (i.e., those in the range 76–100).

APPENDIX A
MARYLAND STATE DEPARTMENT OF EDUCATION
1993 MSPAP STANDARDS REPORT

School System:
School:

| | Standard: Percent EXSAT | | Number Tested | Number Absent/ Excused | Number Not Reported** | Percent at EX* | Percent at SAT* |
|---|---|---|---|---|---|---|---|
| Grade 05 | | | | | | | |
| Reading | 25 | 70 | 60 | 3 | 0 | 4.8 | 28.6 |
| Mathematics | 25 | 70 | 61 | 2 | 0 | 9.5 | 57.1 |
| Science | 25 | 70 | 59 | 4 | 0 | 4.8 | 47.6 |
| Social Studies | 25 | 70 | 60 | 3 | 0 | 1.6 | 38.1 |

* Excellent (EX) and Satisfactory (SAT) are defined as percentages based on the number tested plus the number absent/excused.

** Includes exemptions for special education, LEP, and second-semester transfers.

## APPENDIX B
## MARYLAND STATE DEPARTMENT OF EDUCATION
## 1993 MSPAP PROFICIENCY LEVEL AND PARTICIPATION REPORT

School System:
School:
Grade Level: 05 June Membership: 64

| **Proficiency Level (1=Highest)** | **Number of Students** | **Percent of Test Takers** | **Participation in Assessment** | **Number of Students** | **Percent of June 1993 Membership** |
|---|---|---|---|---|---|
| **READING** | | | | | |
| Level 1 | | | Completed Test | 60 | 93.8 |
| Level 2 | 3 | 5.0 | Absent/Excused | 3 | 4.7 |
| Level 3 | 15 | 25.0 | Exempted* | 0 | 0.0 |
| Level 4 | 31 | 51.7 | 2nd-Sem. Transfers | 0 | 0.0 |
| Level 5 | 11 | 18.3 | Total Students Accounted For | 63 | 98.4 |
| **WRITING** | | | Completed Test | 60 | 93.8 |
| Level ½ | 11 | 18.3 | Absent/Excused | 3 | 4.7 |
| Level 3 | 23 | 38.3 | Exempted | 0 | 0.0 |
| Level 4 | 4 | 6.7 | 2nd-Sem. Transfers | 0 | 0.0 |
| Level 5 | 22 | 36.7 | Total Students Accounted For | 63 | 98.4 |
| **LANGUAGE USAGE** | | | | | |
| Level 1 | 3 | 5.1 | Completed Test | 59 | 92.2 |
| Level 2 | 8 | 13.6 | Absent/Excused | 4 | 6.3 |
| Level 3 | 15 | 25.4 | Exempted | 0 | 0.0 |
| Level 4/5 | 33 | 55.9 | 2nd -Sem. Transfers | 0 | 0.0 |
| | | | Total Students Accounted For | 63 | 98.4 |

** Includes special education students who were exempted only from the reading portion of the assessment.*

*Note: Proficiency levels are defined as percentages based on the number who completed the test.*

BEST COPY AVAILABLE

## APPENDIX C
## MARYLAND STATE DEPARTMENT OF EDUCATION
## 1993 MSPAP PROFICIENCY LEVEL AND PARTICIPATION REPORT

School System:
School:
Grade Level: 05 June Membership: 64

| Proficiency Level (1=Highest) | Number of Students | Percent of Test Takers | Participation in Assessment | Number of Students | Percent of June 1993 Membership |
|---|---|---|---|---|---|
| **MATHEMATICS** | | | | | |
| Level 1 | | | Completed Test | 61 | 95.3 |
| Level 2 | 6 | 9.8 | Absent/Excused | 2 | 3.1 |
| Level 3 | 30 | 49.2 | Exempted* | 0 | 0.0 |
| Level 4 | 23 | 37.7 | 2nd-Sem. Transfers | 0 | 0.0 |
| Level 5 | 2 | 3.3 | Total Students Accounted For | 63 | 98.4 |
| **SCIENCE** | | | | | |
| | | | Completed Test | 59 | 92.2 |
| Level ½ | 3 | 5.1 | Absent/Excused | 4 | 6.3 |
| Level 3 | 27 | 45.8 | Exempted | 0 | 0.0 |
| Level 4 | 27 | 45.8 | 2nd-Sem. Transfers | 0 | 0.0 |
| Level 5 | 2 | 3.4 | Total Students Accounted For | 63 | 98.4 |
| **SOCIAL STUDIES** | | | | | |
| Level 1 | | | Completed Test | 60 | 93.8 |
| Level 2 | 1 | 1.7 | Absent/Excused | 3 | 4.7 |
| Level 3 | 23 | 38.3 | Exempted | 0 | 0.0 |
| Level 4/5 | 36 | 60.0 | 2nd-Sem. Transfers | 0 | 0.0 |
| | | | Total Students Accounted For | 63 | 98.4 |

* Includes special education students who were exempted only from the reading portion of the assessment.

Note: Proficiency levels are defined as percentages based on the number who completed the test.

BEST COPY AVAILABLE

# APPENDIX D
# MARYLAND STATE DEPARTMENT OF EDUCATION
# 1993 MSPAP DISAGGREGATED PROFICIENCY LEVEL REPORT

School System:
School:
Grade Level: 05
MATHEMATICS

**Percentage of Students At:**

| | Number | Level 1 | Level 2 | Level 3 | Level 4 | Level 5 |
|---|---|---|---|---|---|---|
| American Indian/ Alaska Native | | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Male | | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Female | | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Asian/ Pacific Islander | | 0.0 | 0.0 | 50.0 | 50.0 | 0.0 |
| Male | * | 0.0 | 0.0 | 50.0 | 50.0 | 0.0 |
| Female | | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| African American (Not Hispanic) | | 0.0 | 0.0 | 55.6 | 37.0 | 7.4 |
| Male | * | 0.0 | 0.0 | 61.5 | 38.5 | 0.0 |
| Female | * | 0.0 | 0.0 | 50.0 | 35.7 | 14.3 |
| White (Not Hispanic) | | 0.0 | 19.4 | 45.2 | 35.5 | 0.0 |
| Male | * | 0.0 | 23.1 | 30.8 | 46.2 | 0.0 |
| Female | * | 0.0 | 16.7 | 55.6 | 27.8 | 0.0 |
| Hispanic | | 0.0 | 0.0 | 0.0 | 100.0 | 0.0 |
| Male | * | 0.0 | 0.0 | 0.0 | 100.0 | 0.0 |
| Female | | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| All Groups | | 0.0 | 9.8 | 49.2 | 37.7 | 3.3 |
| Male | | 0.0 | 10.3 | 44.8 | 44.8 | 0.0 |
| Female | | 0.0 | 9.4 | 53.1 | 31.3 | 6.3 |

* Fewer than 20 students

Note: Proficiency levels are defined as percentages based on the number who completed the test.

## APPENDIX E
## MARYLAND STATE DEPARTMENT OF EDUCATION
## 1993 MSPAP PROFICIENCY LEVEL AND PARTICIPATION REPORT

School System:
School:
Grade Level: 05

| | | | | | | | | PERCENTAGE WITH OUTCOME SCORES IN: | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | 0-25 | | 26-50 | | 51-75 | | 76-100 |
| | DIFF | MEAN | MDN | SD | HI | LOW | TOTAL | N | % | N | % | N | % | N | % |
| READING | | | | | | | | | | | | | | | |
| 1. For Literary Experience | 27 | 37.42 | 30.0 | 23.36 | 87 | 4 | 19 | 5 | 26.3 | 10 | 52.6 | 1 | 5.3 | 3 | 15.8 |
| 2. To be Informed | 27 | 26.52 | 25.0 | 19.88 | 75 | 1 | 60 | 34 | 56.7 | 19 | 31.7 | 7 | 11.7 | 0 | 0.0 |
| 3. To Perform a Task | 36 | 45.45 | 43.0 | 19.65 | 86 | 1 | 20 | 2 | 10.0 | 12 | 60.0 | 5 | 25.0 | 1 | 5.0 |
| WRITING | | | | | | | | | | | | | | | |
| 1. To Inform | 24 | 39.64 | 33.0 | 29.88 | 100 | 4 | 39 | 13 | 33.3 | 16 | 41.0 | 6 | 15.4 | 4 | 10.3 |
| 2. To Persuade | 30 | 28.85 | 16.0 | 22.81 | 65 | 9 | 41 | 24 | 58.5 | 3 | 7.3 | 14 | 34.1 | 0 | 0.0 |
| 3. To Express Personal Ideas | 26 | 52.95 | 33.0 | 28.57 | 100 | 17 | 19 | 1 | 5.3 | 10 | 52.6 | 4 | 21.1 | 4 | 21.1 |
| LANGUAGE USAGE | | | | | | | | | | | | | | | |
| 1. Language in Use | 33 | 46.10 | 41.0 | 25.68 | 100 | 3 | 59 | 14 | 23.7 | 20 | 33.9 | 14 | 23.7 | 11 | 18.6 |
| MATHEMATICS | | | | | | | | | | | | | | | |
| 1. Problem Solving | 33 | 57.25 | 79.0 | 46.88 | 100 | 0 | 40 | 15 | 37.5 | 0 | 0.0 | 5 | 12.5 | 20 | 50.0 |
| 2. Communication | 29 | 38.59 | 40.0 | 20.53 | 87 | 3 | 61 | 19 | 31.1 | 26 | 42.6 | 12 | 19.7 | 4 | 6.6 |
| 3. Reasoning | 31 | 32.84 | 36.0 | 21.74 | 85 | 4 | 61 | 20 | 32.8 | 27 | 44.3 | 13 | 21.3 | 1 | 1.6 |
| 4. Connections | 22 | 30.03 | 17.0 | 34.49 | 100 | 1 | 61 | 34 | 55.7 | 11 | 18.0 | 8 | 13.1 | 8 | 13.1 |
| 5. Estimation | 0 | | | | | | 0 | | | | | | | | |
| 6. Arithmetic Operations | 34 | 47.02 | 46.0 | 18.87 | 88 | 6 | 61 | 9 | 14.8 | 24 | 39.3 | 25 | 41.0 | 3 | 4.9 |
| 7. Number Relationships | 27 | 46.60 | 43.0 | 21.23 | 77 | 4 | 40 | 8 | 20.0 | 14 | 35.0 | 14 | 35.0 | 4 | 10.0 |
| 8. Geometry | 39 | 54.21 | 54.0 | 27.04 | 100 | 7 | 61 | 4 | 6.6 | 24 | 39.3 | 20 | 32.8 | 13 | 21.3 |
| 9. Measurement | 33 | 53.51 | 58.0 | 23.23 | 100 | 3 | 61 | 8 | 13.1 | 18 | 29.5 | 27 | 44.3 | 8 | 13.1 |
| 10. Statistics | 29 | 40.03 | 42.0 | 24.98 | 100 | 4 | 61 | 19 | 31.1 | 22 | 36.1 | 16 | 26.2 | 4 | 6.6 |
| 11. Probability | 27 | 41.05 | 42.0 | 27.04 | 100 | 2 | 40 | 12 | 30.0 | 15 | 37.5 | 8 | 20.0 | 5 | 12.5 |
| 12. Patterns/Relationships | 40 | 47.50 | 42.0 | 23.18 | 100 | 15 | 40 | 7 | 17.5 | 18 | 45.0 | 10 | 25.0 | 5 | 12.5 |
| 13. Algebra | 30 | 41.65 | 34.0 | 29.50 | 100 | 3 | 40 | 12 | 30.0 | 9 | 22.5 | 14 | 35.0 | 5 | 12.5 |
| SCIENCE | | | | | | | | | | | | | | | |
| 1. Concepts of Science | 35 | 40.75 | 44.0 | 17.24 | 85 | 5 | 59 | 10 | 16.9 | 35 | 59.3 | 12 | 20.3 | 2 | 3.4 |
| 2. Nature of Science | 36 | 41.81 | 44.0 | 15.62 | 80 | 5 | 59 | 11 | 18.6 | 35 | 59.3 | 10 | 16.9 | 3 | 5.1 |
| 3. Habits of Mind | 36 | 45.15 | 47.0 | 20.29 | 89 | 5 | 59 | 14 | 23.7 | 24 | 40.7 | 18 | 30.5 | 3 | 5.1 |
| 4. Processes of Science | 32 | 42.98 | 41.0 | 12.61 | 83 | 11 | 59 | 1 | 1.7 | 41 | 69.5 | 16 | 27.1 | 1 | 1.7 |
| 5. Applications of Science | 32 | 42.46 | 39.0 | 22.01 | 100 | 2 | 59 | 11 | 18.6 | 29 | 49.2 | 14 | 23.7 | 5 | 8.5 |
| SOCIAL STUDIES | | | | | | | | | | | | | | | |
| 1. Political Systems | 17 | 21.66 | 23.0 | 14.75 | 52 | 2 | 41 | 25 | 61.0 | 14 | 34.1 | 2 | 4.9 | 0 | 0.0 |
| 2. Peoples/Nation & World | 24 | 27.63 | 20.0 | 18.46 | 62 | 3 | 40 | 21 | 52.5 | 15 | 37.5 | 4 | 10.0 | 0 | 0.0 |
| 3. Geography | 36 | 43.62 | 41.5 | 22.04 | 100 | 5 | 60 | 14 | 23.3 | 26 | 43.3 | 16 | 26.7 | 4 | 6.7 |
| 4. Economics | 19 | 17.97 | 17.0 | 14.56 | 64 | 3 | 39 | 29 | 74.4 | 8 | 20.5 | 2 | 5.1 | 0 | 0.0 |
| 5. Skills and Processes | 27 | 31.95 | 33.0 | 13.88 | 68 | 5 | 60 | 20 | 33.3 | 36 | 60.0 | 4 | 6.7 | 0 | 0.0 |
| 6. Valuing Self and Others | 21 | 20.97 | 22.0 | 15.48 | 67 | 3 | 60 | 38 | 63.3 | 19 | 31.7 | 3 | 5.0 | 0 | 0.0 |
| 7. Understand./Attitudes | 20 | 23.78 | 23.5 | 18.73 | 69 | 3 | 60 | 36 | 60.0 | 20 | 33.3 | 4 | 6.7 | 0 | 0.0 |

# VII-10

# Improving Assessment and Reporting in the Edmonton Public Schools

DALE ARMSTRONG
EDMONTON PUBLIC SCHOOLS, EDMONTON, ALBERTA, CANADA

When my colleagues and I set out to improve large-scale student assessment in Edmonton Public Schools, we identified the first priority to be a review of existing indicators and measures of district performance, including the results of the following:

- Annual districtwide surveys of staff and student attitudes
- Biennial surveys of parent and general community attitudes
- Districtwide achievement tests for grades 3, 6, and 9 in mathematics, language arts, science, and social studies
- Alberta Education achievement tests for grades 3, 6, and 9
- Alberta Education diploma examinations for grade 12
- International Baccalaureate examinations

Although these indicators and measures provided useful information about the district's overall performance, there were obvious deficiencies in student achievement strategies and measures. For instance, the curriculum-based tests were unable to differentiate among grade levels of achievement except at the 3, 6, 9, and 12 levels. This type of externally evaluated paper-and-pencil testing at widely dispersed grade levels was inadequate. Moreover, the tests were unable to extract from these results any measures of student performance related to the district's six essential learning outcomes: communication, responsible citizenship well-being, knowledge, inquiry, and aesthetic appreciation.

To achieve the goal of providing students, parents, staff, the board, and the general community with valid, accurate, and useful information about student achievement and growth, we identified the following project needs:

- Instruments and strategies that will facilitate the assessment of student achievement and the annual communication of assessment results while satisfying the requirements of the district, the school, and individual students.
- Standards against which student achievement is evaluated relative to the graded curriculum.
- Assessment, evaluation, and communication of students' continuous development of skills, attitudes, and knowledge.

- Strategies for gathering teachers' judgments and other evidence of students' mastery of the curriculum relative to the essential learning outcomes.
- Opportunity for students to demonstrate their highest level of achievement.

## Compelling Reasons for Change

> The assessments that count—the assessments that most strongly influence student learning and academic self-concept—are those developed and used by teachers in the classroom.
>
> — Richard J. Stiggins, "Revitalizing Classroom Assessment: The Highest Instructional Priority." *Phi Delta Kappan* 69, 5 (January 1988): 363-368.

Our improvement effort was influenced by changes in other parts of the education system. First, the Edmonton Public Schools had to meet legal requirements established by Section 13 of the *School Act*, specifically: "The student's teacher is responsible for assessing and periodically reporting student achievement."

The staff also developed certain assumptions that we wanted our assessment process to support:

- The achievement and performance of each student are measurable.
- Student growth in achievement is one of the indicators for measuring student, school, and district performance.
- Teacher judgment about student achievement and performance is valid and reliable.

We also wanted our assessment system to be guided by specific principles agreed upon by the assessment branch of the Edmonton Public Schools:

- Assessment, evaluation, and communication of student growth are based on the knowledge, skills, and attitudes students are expected to learn. The methods of assessment and evaluation vary and are consistent with established principles of learning and instruction. Information about methods of assessment and results of evaluation is communicated to students, parents, and the community.
- Student growth is assessed, evaluated, and communicated for all outcomes.
- Student growth is evaluated and communicated on an ongoing basis and is used to plan programming.
- Student growth is demonstrated through a variety of performances evaluated by the teacher.
- Student growth is enhanced when students participate in the assessment-evaluation-communication process.
- Student growth is enhanced when assessment, evaluation, and communication are viewed positively by the student.

## Project Design

The framework for the project assumed four major components:

1. Design and validation of instruments and strategies to allow for student demonstration of achievement.
2. Design and validation of strategies and criteria to allow for teacher judgment, data collection, and confirmation.
3. Investigation of workable applications of the "portfolio assessment" concept.
4. Development of prototype technology that would accommodate school and district needs for computer display, communication, and storage of information about student achievement and growth.
5. Overall management of the project to ensure coordination of the four development areas.

## Measuring Growth

To enhance current district and provincial assessment practices and to overcome the deficiencies of conventional paper-and-pencil "fill-in-the-blank" or "circle-the-correct-answer" assessment instruments, we decided that assessments should emphasize the concept of multiple activity tasks linked to a set of essential learning outcomes. We established the following requirements for the design of an initial task for field testing:

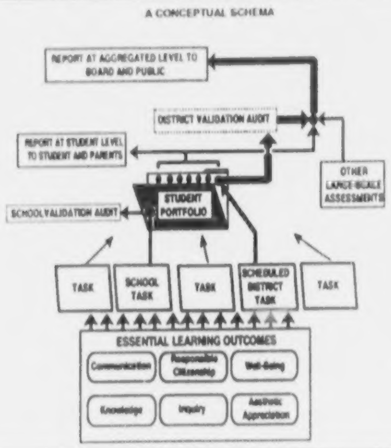
- Tasks are seen to be valid and reliable by students, staff, parents, and the community, and these publics are satisfied with the information they are supplied about student growth.
- The tasks are based on essential learning outcomes.
- Teachers are involved in the development of criteria and the growth measure.
- Criteria are developed to evaluate the grade level of achievement.
- The tasks are undertaken by students currently demonstrating a wide range of performance and achievement.
- The tasks are as typical as possible of normal classroom and real-life experiences and involve the students in multiple activities.
- There is opportunity for student-to-student and teacher-to-student interactions.
- Classroom teachers are trained to use student products in making judgments about student achievement.

### Portfolio Assessment

To guide the district's developmental work on portfolio assessment, the Edmonton Public Schools have defined a student portfolio as a purposeful collection of student products that exhibits to the student and others the student's demonstration of achievement, growth, and performance in essential learning outcomes. This collection includes criteria for selection, achievement criteria, student self-reflections, and teacher notes describing the context within which the products were obtained. The student is involved in selecting the contents. Figure 1 shows a conceptual schema of the relationship of student portfolios to an enhanced school and district assessment system.

FIGURE 1

PORTFOLIO ASSESSMENT SYSTEM



The relationship of the portfolio technique to developmental assessment is aptly described by David Elkind:

> Developmental assessment involves documenting the work that a child has done over a period of time. Usually this is done by having a child keep a portfolio that includes all of his or her writing, drawing, math, explorations and so on. In looking through such a portfolio, we can get a good idea of the quality of work that the child is capable of doing and of his or her progress over the given period(Early Education Conference Presentation, Edmonton, 1989).

The initial developmental work on the student portfolio system focused on the following issues:

- Application—usefulness of the information
- Responsibility—roles of students, teachers, parents, community, administrators, and trustees
- Content—what is developed and how
- Design—task development, validation, and administration
- Evaluation—criteria for making judgments
- Management and logistics—storage, movement, access, *School Act* requirements for student records
- Implementation—staff development and training
- Reporting—communication of student growth information to the many publics.

The development of workable strategies for teachers' use of portfolio assessment in the day-to-day realities of the classroom took priority over any concerns about the physical format of a portfolio. We had to differentiate between the use of portfolio contents for individual student assessment and for district accountability..

Teachers have been very involved in developing appropriate teaching strategies and in identifying a realistic set of student products to include in the portfolio. They will be looking at the following products to decide whether they are suitable for the student portfolio:

- Narrative, personal, and informational writing
- Students' own self-reflection analyses
- Photographs, videocassettes, and audiotapes
- Letters of reference
- Tables and reports of investigative student work
- Models
- Artwork of all kinds
- Student-generated computer disks
- Teacher judgments, assessment criteria, task descriptions, exemplars, and context notes.

### Prototype Computer Technology

In the initial project design, we identified two major logistical restraints to obtaining maximum benefit from the proposed assessment process: (1) the human limitations of a teacher coping with a wide array of media and achievement criteria; and (2) the difficulties of communicating both aggregated information about school and student performance (to the community) and information about individual student growth (to parents and teachers). The design therefore included the following objectives:

- To investigate the potential of current "pen-based" computer technology.
- To develop prototype user screens for entering and presenting information about students' growth.
- To identify requirements for upgrading the district's wide area computer network to accommodate the bandwidth required for the central collection of student achievement data, including compressed image data transfer.
- To develop computer applications that teachers can use to prepare meaningful progress reports for parents.

BEST COPY AVAILABLE

## Developing New Assessment Strategies

> To solely use standardized achievement tests is like casting a net into the sea—a net that is intentionally designed to let the most interesting fish get away. Then, to describe the ones that are caught strictly in terms of their weight is to radically reduce what we know about them. To further conclude that all the contents of the sea consist of fish like those in the net compounds the error further. We need to know more about those we catch. We need new nets.
>
> — Dr. William T. Randall, Commissioner of Education, State of Colorado (ASCD Consortium Presentation, Colorado, 1992)

The new assessment strategies designed, field-tested, and piloted in this development project are supplements to, rather than replacements for, the traditional and current assessment practices. The results demonstrate the potential for the *enhancement* of existing student assessments in our schools. Figure 2 summarizes some of the changes we have made.

### Field Testing Prototype Tasks

In June 1990, we began the process of field testing assessment tasks with our students. Figure 3 presents a chronology of these efforts, which began with the field-test of a prototype task with "year-one" students (i.e., students in their first year of schooling, excluding kindergarten) in nine elementary schools. This multiple-activity task was based on the general theme of "change" and required students to participate in at least two writing activities as well as drawing, self-reflection, and reading activities.

Teachers judged students' achievement level using an initial set of achievement criteria. The field test addressed essential learning outcomes, indicators, grade level of achievement, and identification of the student products being evaluated. The essential learning outcomes and selected indicators are shown in Figure 4.

The results of the June 1990 prototype task were analyzed to develop achievement criteria and a more refined task for use in the 1990–91 school year. Committees of teachers, district consultants, and curriculum specialists carried out this analysis and development.

Inservice training was provided in May 1991 to teachers of year-one students in the nine schools selected to pilot the more comprehensive measure to be administered that month (administration to other students was optional). Students' work was collected for analysis and central scoring in July 1991.

During the 1991–92 school year, we conducted trials of 15 tasks developed for students in years 1 through 12 at 35 demonstration sites (Figure 5 lists these tasks); we also included the students now in year 2 at the original nine pilot schools. To maintain the reliability of procedures and results, teachers at demonstration sites were provided with a specific administration routine for each task, including a statement of purpose, a process model, and an outline of steps and activities.

FIGURE 2
COMPARISON OF ASSESSMENT METHODOLOGIES

| **"Old Nets"** | **"New Nets"** |
|---|---|
| Assess student achievement after 3, 6, and 9 years of schooling | Assess student achievement after each year of schooling |
| Assess student achievement relative to curriculum for grades 3, 6, and 9 | Assess student achievement relative to all levels of the curriculum |
| Assess student achievement relative to language arts, mathematics, science, and social studies | Assess student achievement relative to the total curriculum |
| Provide school and district information about student achievement at four points in the student's career | Annually provide school and district information about student growth in achievement |
| Assessment is external to instruction—marked by someone other than the student's teacher | Assessment is part of instruction—determined by the student's teacher—subject to validation and audit at the district level |
| Assessment is based on student responses on a "paper-and- pencil" measure | Assessment is based on student performance of assigned tasks including but not limited to "paper-and-pencil" measures |
| Analysis of each student's performance on four tests is made and results are used by the teacher | Each student's performance on tasks is used by teacher, student, and parent for diagnosis, planning, and celebration of success |

Analysis of the results of the 1991–92 assessment tasks continued during the 1992–93 school year, in preparation for the administration of the tasks as growth measures in April 1993. Figure 6 shows the results of the tasks piloted in the 1991–92 school year. All student work from the tasks was judged separately by the student's teacher and by analysts who had had major involvement in the establishment of scoring criteria. The levels of agreement on each indicator are shown in Figure 7.

**Development of Achievement Criteria.** Achievement criteria for selected indicators of communication (the first learning outcome) for grades 1 through 12 were developed in 1990–91 by a committee of ten teachers and two consultants. These indicators were revised and refined in July 1991 for grade 1, and during 1991–92 for grade 2. Preliminary criteria for selected indicators of inquiry (the fifth learning outcome) for grades 1 through 12 were developed during 1991–92 and revised and refined during 1992–93. Preliminary criteria for the selected indicators of the other four essential learning outcomes were developed during 1992–93.

**Curricular Relevance.** Developmental work has proceeded on the topic of relating essential learning outcomes and selected indicators to language arts, mathematics, science, social studies, health, and art.

**District Implications.** Multiple strategies and measures for student assessment are now included in district practices in the Edmonton Public Schools. Current and proposed practices in each of the 199 schools are guided by the ten principles for assessment, evaluation, and communication of student growth.

## Additional Developments

### Portfolio Design

Initial work on establishing suitable physical designs for a student portfolio was an important catalyst for the examination of more vital issues, such as the possible range of contents in a portfolio, the implications of regulations under the *School Act* that define contents of a student record and the access to that record, and logistical problems that could be associated with full use of a portfolio assessment system. Models of suitable portfolio "containers" were produced, but it soon became clear that the most suitable physical arrangements for managing most of the potential contents of a student's portfolio were those developed at the local school and classroom level. The number of methods used is likely to be as numerous as the number of teachers employing portfolio assessment procedures.

### Prototype Computer Technology I: Pen-Based Systems

The two competing devices on the market were evaluated for use as data collection systems used by teachers in the classroom. Both computers showed instability in their operating systems. Hand-printed character recognition was inconsistent and unreliable. Acquisition costs were high, and the devices were too bulky for in a classroom setting. No further evaluation or development of suitable application software is contemplated at this time.

FIGURE 3

CHRONOLOGY OF ASSESSMENT TASK DEVELOPMENT AND FIELD TESTING

1989–90
- One task developed and field-tested in nine elementary schools

1990–91
- Prototype task developed and field-tested in nine elementary schools at the one-year level
- Achievement criteria established and validated by teacher, consultant, and curriculum specialist teams

1991–92
- 14 tasks developed and the 1989–91 tasks revised
- 15 tasks piloted in 35 schools representing "years in school" 1 to 12 (see Figure 5)
- Each task cross-referenced to essential learning outcomes and relevant course outcomes and expectations
- Inservice training and consultant help to complete judgments for each essential learning outcome and selected indicators provided to teachers
- Achievement information recorded by teachers
- More than 2,000 students participate in the administration of the tasks

1992–93
- Analysis of the teacher judgments against criteria provides initial verification of the validity of the assessment instruments and strategies

FIGURE 4

LEARNING OUTCOMES AND SELECTED INDICATORS

| **Learning Outcome** | **Selected Indicators** |
|---|---|
| COMMUNICATION | reading<br>writing |
| RESPONSIBLE CITIZENSHIP | participating<br>cooperating<br>providing service |
| WELL-BEING | making choices<br>exhibiting positive life habits |
| KNOWLEDGE | applying<br>calculating |
| INQUIRY | hypothesizing<br>researching<br>decision making |
| AESTHETIC APPRECIATION | valuing<br>expressing |

BEST COPY AVAILABLE

FIGURE 5

ASSESSMENT TASKS AND ESSENTIAL LEARNING OUTCOMES

| No. | Task Name | Essential Learning Outcomes Demonstrated* |
|---|---|---|
| 1 | Patterns | All 6 outcomes |
| 2 | Relationships | All 6 outcomes |
| 3 | Adapting to the Environment | All 6 outcomes |
| 4 | Opposites | All 6 outcomes |
| 5 | Traffic Flow | All 6 outcomes |
| 6 | Points of View | Communication |
| 7 | Linking the Past to Today | Communication |
| 8 | Heroes and Heroines | Communication |
| 9 | Literary Traditions | Communication |
| 10 | A Greek Myth | Communication |
| 11 | Limits to Population | Inquiry |
| 12 | "Roto-copters" | Inquiry |
| 13 | Exploration of Shapes | Inquiry |
| 14 | Transportation | Inquiry |
| 15 | All Summer In A Day | Inquiry |

* Essential Learning Outcomes

1. Communication
2. Responsible Citizenship
3. Well-being
4. Knowledge
5. Inquiry
6. Aesthetic Appreciation

### Prototype Computer Technology II: User Screens

A series of prototype screens has been developed using the standard "Windows" screen format and Visual Basic as the developmental language. Development is continuing parallel to the development of assessment tasks and related reporting procedures in the design project. The flexibility of the design tool is such that suitable screens and the underlying application for the capture and presentation of information on student growth can now be quickly developed, including the incorporation of scanned image and recorded sound when appropriate. Figures 8 through 12 show screen images from the various interfaces developed to date.

### Prototype Computer Technology III: Requirements for Wide Area Network Upgrading

The technical requirements for an upgraded wide area network to serve over 200 remote locations were identified, and the system was implemented in September 1994. It addresses the need to accommodate multiple simultaneous users, high data traffic flow, image transfer, distributed processing of data where appropriate, and internal school and central connections to local area networks. The recent developments in data communications standards and more affordable technology have made the technical concept feasible. The wide area network, which connects all schools and central operating units independent of computer type, meets the data communication needs and storage demands of the district's monitoring functions. Issues that continue to be addressed include:

- Cooperative line charge arrangements with the carriers
- Continuing instability of local area networks
- Security of data
- The role of "multimedia" in the assessment process and the technical challenge of handling heavy data traffic on a wide area network
- Improvements in client-server design and accommodation of complex calls to a relational database.

### Prototype Computer Technology IV: Computer-Assisted Student Progress Reporting

A flexible and stable application has been developed to allow teachers to key in their own commentary for each student and produce a progress report, via local area network, that prints in a central location in the school office. The report includes the student's picture, comments, and "marks." The system is in use and incorporates many design features suggested by classroom teachers.

There is a clear connection between this developmental work and the work related to user screens for the presentation and capture of information about student growth. The current progress reporting system is designed for the Macintosh computer, but a DOS-Windows version will shortly be available for use by a wider range of schools. The system is compatible with wide area network standards. We anticipate that the two initiatives (progress reporting and student growth user screens) will be merged.

In addition, encouraging progress is being made in the development of an application that will allow teachers to access central student records and achievement history. The application is designed to be used on DOS-Windows or Macintosh platforms and will ensure the integrity of data regarding students and their achievement.

## Conclusions

Much follow-up work awaits us:

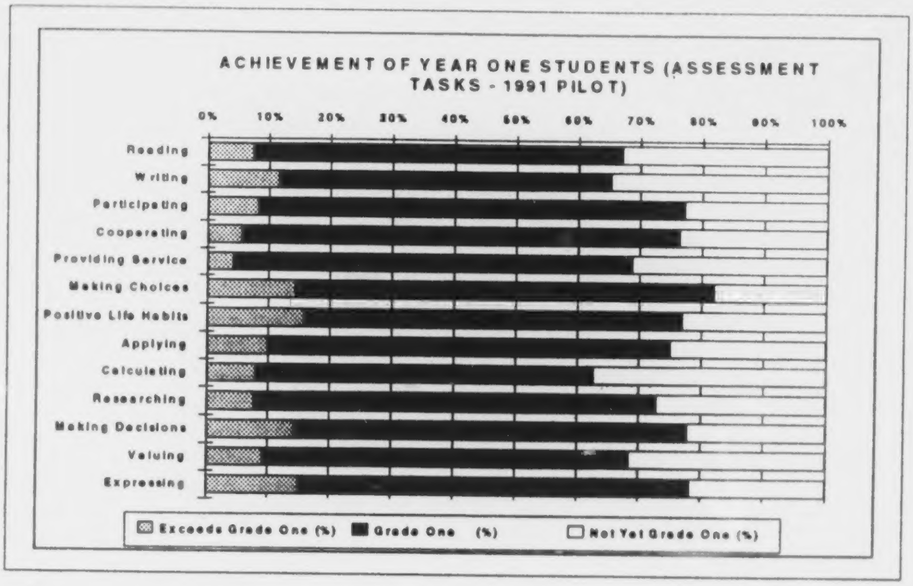
- Revision and design of tasks for students in years one to twelve
- Preparation of teachers for task administration
- Administration of revised and developed tasks in spring 1994
- Refinement, development, and application of achievement criteria
- Judgment of student achievement in the demonstration schools
- Collection and reporting of growth information
- Development of paper and computer formats for display of growth data at all levels
- Preparation of a staff development plan
- Attempt to resolve identified issues

Although we have many unanswered questions, the findings from the first three years of the project are encouraging and give promise for the future (see Figures 13 and 14). Comments from teachers and students indicate that the tasks and results can be useful in enhancing the reporting of student achievement and growth. The data derived from the teacher evaluations can be used by teachers as part of their assessment of a student's achievement and growth and can be aggregated at the school and district levels to provide useful accountability information. At the same time, portfolios of student work can be used during parent-teacher-student conferences to discuss and validate a student's individual progress.

The feasibility of developing easy-to-use computer applications for the capture, display, and communication of student achievement and growth has been established, and applications to meet this objective have been developed.

FIGURE 6

ACHIEVEMENT RESULTS OF ASSESSMENT TASKS PILOTED IN 1991–92



BEST COPY AVAILABLE

FIGURE 7

CORRELATION BETWEEN GRADING JUDGMENTS OF ANALYSTS AND TEACHERS

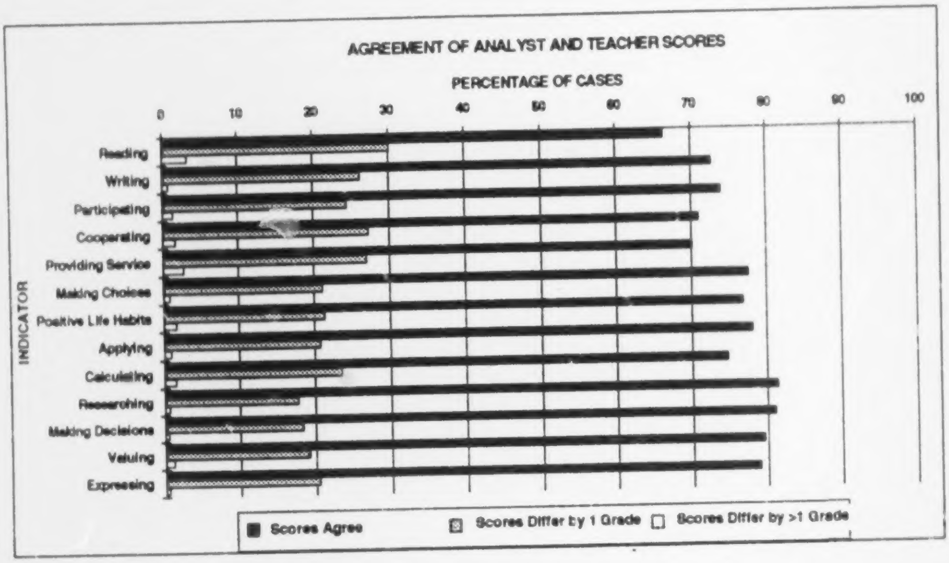
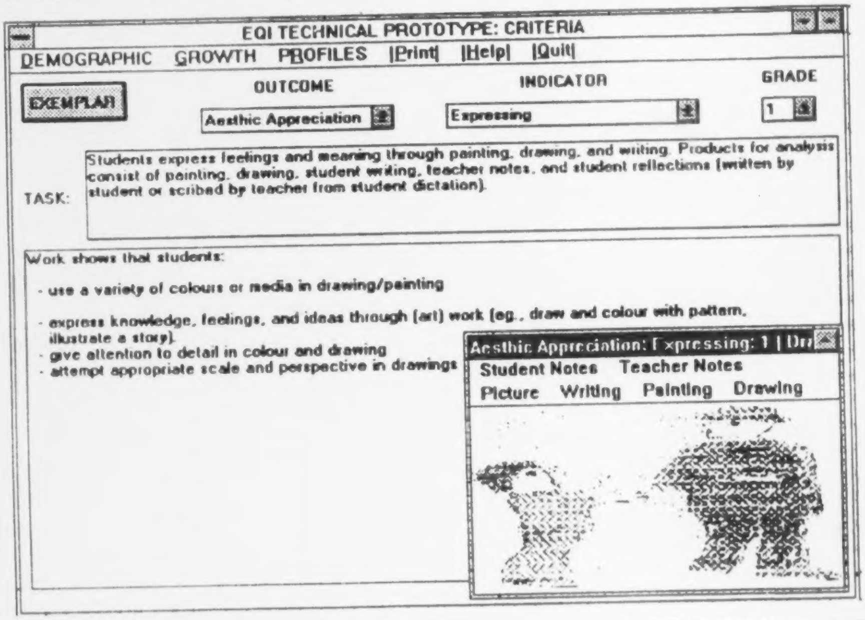


FIGURE 8

A SCREEN TO DISPLAY CRITERIA AND A SCANNED IMAGE OF STUDENT WORK


EQI TECHNICAL PROTOTYPE: CRITERIA
DEMOGRAPHIC
GROWTH
PROFILES
|Print|
|Help|
|Quit|
EXEMPLAR
Aesthic Appreciation: Expressing: 1 | Dr
Student Notes
Teacher Notes
Picture
Writing
Painting
Drawing

OUTCOME: Aesthic Appreciation

INDICATOR: Expressing

GRADE: 1

TASK: Students express feelings and meaning through painting, drawing, and writing. Products for analysis consist of painting, drawing, student writing, teacher notes, and student reflections (written by student or scribed by teacher from student dictation).

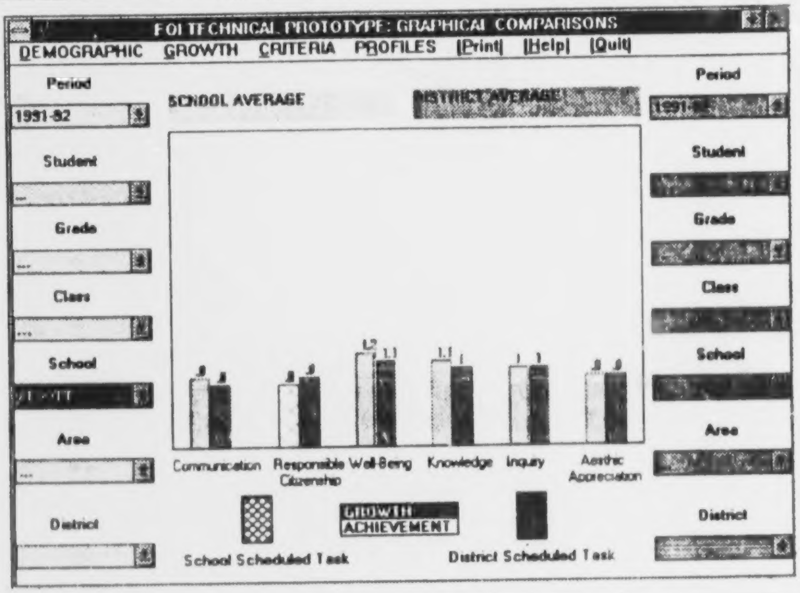
Work shows that students:

- use a variety of colours or media in drawing/painting
- express knowledge, feelings, and ideas through (art) work (eg., draw and colour with pattern, illustrate a story).
- give attention to detail in colour and drawing
- attempt appropriate scale and perspective in drawings

BEST COPY AVAILABLE

FIGURE 9

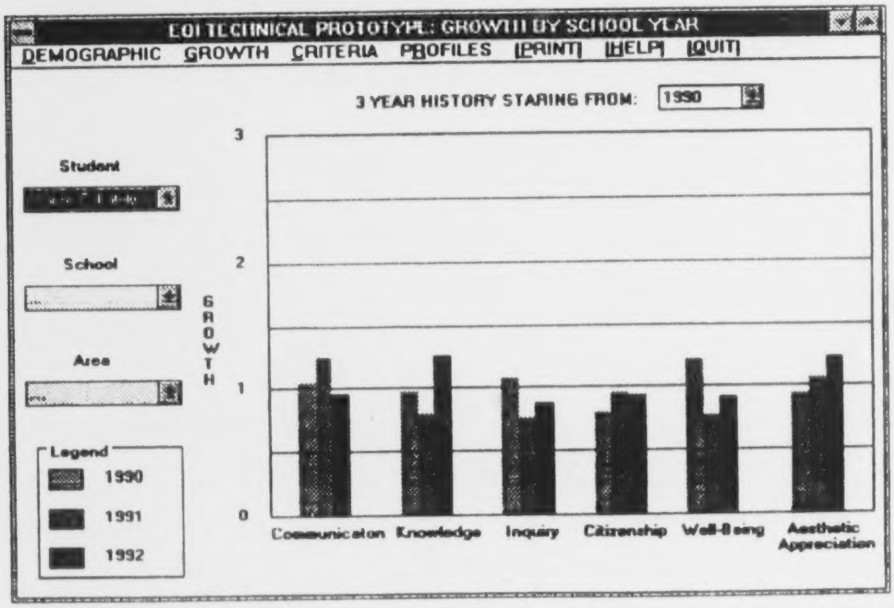
A SCREEN TO DISPLAY GRAPHICAL COMPARISONS AT STUDENT, CLASS, SCHOOL, DISTRICT, OR DATE LEVEL


EOI TECHNICAL PROTOTYPE: GRAPHICAL COMPARISONS
DEMOGRAPHIC
GROWTH
CRITERIA
PROFILES
[Print]
[Help]
[Quit]
Period
Student
Grade
Class
School
Area
District
SCHOOL AVERAGE
DISTRICT AVERAGE
.9
.8
.8
.9
1.2
1.1
1.1
1
1
1
.9
.9
Communication
Responsible Citizenship
Well-Being
Knowledge
Inquiry
Aesthetic Appreciation
GROWTH
ACHIEVEMENT
School Scheduled Task
District Scheduled Task

Period: 1991-92

FIGURE 10

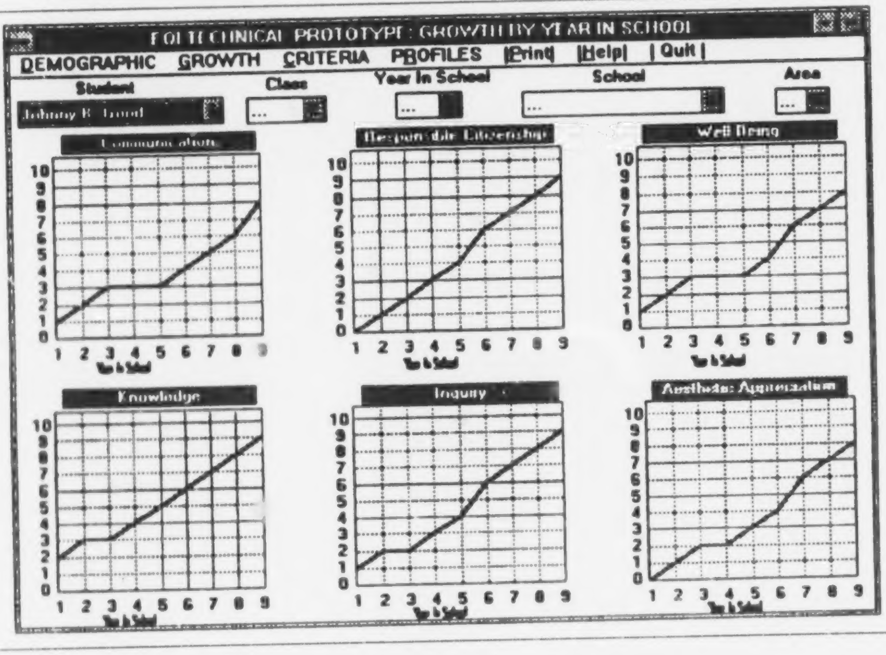
A SCREEN TO DISPLAY STUDENT GROWTH BY SCHOOL YEAR


EOI TECHNICAL PROTOTYPE: GROWTH BY SCHOOL YEAR
DEMOGRAPHIC
GROWTH
CRITERIA
PROFILES
[PRINT]
[HELP]
[QUIT]
Student
School
Area
GROWTH
3
2
1
0
Communication
Knowledge
Inquiry
Citizenship
Well-Being
Aesthetic Appreciation
Legend
1990
1991
1992

3 YEAR HISTORY STARING FROM: 1990

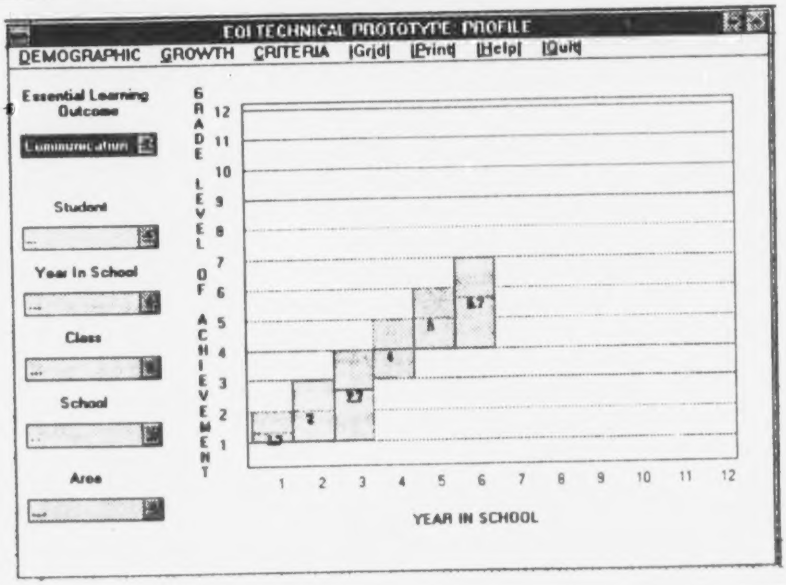
BEST COPY AVAILABLE

Figure 11
A SCREEN TO DISPLAY STUDENT GROWTH BY YEAR IN SCHOOL


EQI TECHNICAL PROTOTYPE: GROWTH BY YEAR IN SCHOOL
DEMOGRAPHIC
GROWTH
CRITERIA
PROFILES
|Print|
|Help|
| Quit |
Communication
Responsible Citizenship
Well Being
Knowledge
Inquiry
Aesthetic Appreciation
0 1 2 3 4 5 6 7 8 9 10
1 2 3 4 5 6 7 8 9
Year in School

Student: Johnny B. Good
Class: ...
Year In School: ...
School: ...
Area: ...

Figure 12
A SCREEN TO DISPLAY PROFILES BY STUDENT, CLASS, SCHOOL, AREA, OR DISTRICT
(BASED ON RANGE OF ACHIEVEMENT BY YEAR IN SCHOOL)


EQI TECHNICAL PROTOTYPE: PROFILE
DEMOGRAPHIC
GROWTH
CRITERIA
|Grid|
|Print|
|Help|
|Quit|
GRADE LEVEL OF ACHIEVEMENT
1 2 3 4 5 6 7 8 9 10 11 12
1.5
2
2.7
4
5
5.7
1 2 3 4 5 6 7 8 9 10 11 12
YEAR IN SCHOOL

Essential Learning Outcome: Communication
Student: ...
Year In School: ...
Class: ...
School: ...
Area: ...

BEST COPY AVAILABLE

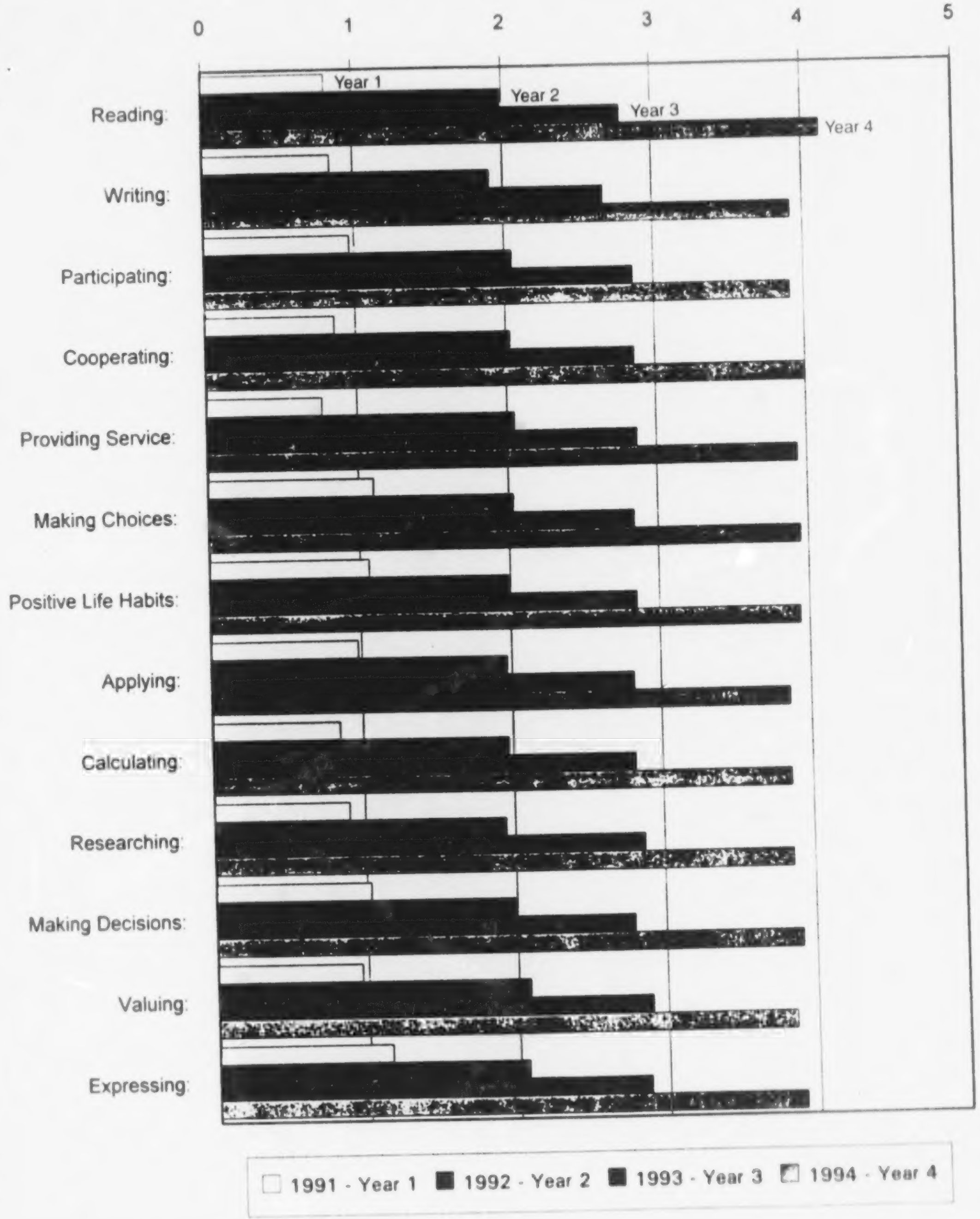
## FIGURE 13
## PERCENTAGE OF STUDENTS AT GIVEN GRADE LEVELS OF ACHIEVEMENT IN ESSENTIAL LEARNING OUTCOMES

| | | Not Yet Grade 1 | Grade 1 | Grade 2 | Grade 3 | Grade 4 | Grade 5 |
|---|---|---|---|---|---|---|---|
| READING | 1991 Year 1 | 33.7 | 48.2 | 18.1 | 0 | 0 | 0 |
| | 1992 Year 2 | 1 | 23.8 | 53.5 | 21.7 | 0 | 0 |
| | 1993 Year 3 | 0 | 2.9 | 28.5 | 55.2 | 3.4 | 0 |
| WRITING | 1991 Year 1 | 42.6 | 41.6 | 15.8 | 0 | 0 | 0 |
| | 1992 Year 2 | 1 | 25.6 | 57.6 | 15.8 | 0 | 0 |
| | 1993 Year 3 | 0 | 3.5 | 31.4 | 59.3 | 5.8 | 0 |
| PARTICIPATING | 1991 Year 1 | 24.6 | 59.5 | 15.9 | 0 | 0 | 0 |
| | 1992 Year 2 | 1.5 | 10.8 | 68.5 | 19.2 | 0 | 0 |
| | 1993 Year 3 | 0 | 2.3 | 19 | 70.1 | 7.5 | 1.1 |
| COOPERATING | 1991 Year 1 | 25.1 | 61.5 | 13.4 | 0 | 0 | 0 |
| | 1992 Year 2 | 1.5 | 10.8 | 70.5 | 17.2 | 0 | 0 |
| | 1993 Year 3 | 0 | 2.3 | 17.8 | 72.4 | 6.3 | 1.1 |
| PROVIDING SERVICE | 1991 Year 1 | 23.2 | 63.7 | 13.2 | 0 | 0 | 0 |
| | 1992 Year 2 | 0.5 | 11.3 | 70.5 | 17.7 | 0 | 0 |
| | 1993 Year 3 | 0 | 1.7 | 18.4 | 72.4 | 6.3 | 1.1 |
| MAKING CHOICES | 1991 Year 1 | 22.4 | 62.7 | 14.9 | 0 | 0 | 0 |
| | 1992 Year 2 | 1 | 8 | 71.1 | 19.9 | 0 | 0 |
| | 1993 Year 3 | 0 | 1.2 | 19.6 | 72.6 | 5.4 | 1.2 |
| POSITIVE LIFE HABITS | 1991 Year 1 | 22.1 | 62.3 | 15.6 | 0 | 0 | 0 |
| | 1992 Year 2 | 1 | 8.5 | 70.6 | 19.9 | 0 | 0 |
| | 1993 Year 3 | 0 | 1.2 | 23.2 | 66.1 | 8.9 | 0.6 |
| APPLYING | 1991 Year 1 | 24.3 | 60.9 | 14.9 | 0 | 0 | 0 |
| | 1992 Year 2 | 1 | 13.7 | 70.1 | 15.2 | 0 | 0 |
| | 1993 Year 3 | 0 | 1.1 | 24.7 | 65.5 | 8.6 | 0 |
| CALCULATING | 1991 Year 1 | 28.2 | 60.4 | 11.3 | 0 | 0 | 0 |
| | 1992 Year 2 | 1 | 13.7 | 71.6 | 13.7 | 0 | 0 |
| | 1993 Year 3 | 0 | 1.7 | 23 | 67.2 | 8 | 0 |
| RESEARCHING | 1991 Year 1 | 27.2 | 61.9 | 10.9 | 0 | 0 | 0 |
| | 1992 Year 2 | 1 | 13.7 | 73.5 | 11.8 | 0 | 0 |
| | 1993 Year 3 | 0 | 2.3 | 17.4 | 71.5 | 8.1 | 0.6 |
| MAKING DECISIONS | 1991 Year 1 | 25.9 | 59.7 | 14.4 | 0 | 0 | 0 |
| | 1992 Year 2 | 0.5 | 13.7 | 72.6 | 13.2 | 0 | 0 |
| | 1993 Year 3 | 0 | 2.9 | 21.1 | 70.2 | 5.8 | 0 |
| VALUING | 1991 Year 1 | 19.1 | 65.7 | 15.2 | 0 | 0 | 0 |
| | 1992 Year 2 | 0 | 8.5 | 71 | 20.5 | 0 | 0 |
| | 1993 Year 3 | 0 | 0.6 | 15.6 | 78 | 5.8 | 0 |
| EXPRESSING | 1991 Year 1 | 15.7 | 63.2 | 21.1 | 0 | 0 | 0 |
| | 1992 Year 2 | 0.5 | 10.5 | 68 | 21 | 0 | 0 |
| | 1993 Year 3 | 0 | 0.6 | 19 | 73 | 7.5 | 0 |

Note: Results based on the performance of 174 students in seven pilot schools.

BEST COPY AVAILABLE

FIGURE 14
STUDENT'S AVERAGE GRADE LEVEL OF ACHIEVEMENT
IN ESSENTIAL LEARNING OUTCOMES



Note: Results based on the performance of 174 students in seven pilot schools.

BEST COPY AVAILABLE

# VIII

# *Implementing Performance Assessment*

# VIII-1

# Implementing Performance Assessment

## Introduction

SHIRLEY M. HORD
SOUTHWEST EDUCATIONAL DEVELOPMENT LABORATORY

GENE E. HALL
UNIVERSITY OF NORTHERN COLORADO

### Why a Section on Change?

Some may ask, "Why have they included a section on implementation and change? Further, why is there a set of papers by authors who know little about performance assessment?" That is because the papers in this section address the other half of the equation—elements of the change process.

As important as each of the first seven sections of this resource book is, and no matter how high the quality of the performance assessments developed, there will be no improvements in schools unless the assessment "innovations" are implemented in classrooms, across schools, and throughout school districts. That means *change*.

Fortunately, there is a knowledge base about the change process that school staffs can use to facilitate implementation of performance assessments. There are practitioners and researchers who have studied the implementation of many educational innovations, as well as the change process in schools more generically. Some of their findings and concepts should be very helpful in planning, implementing, and evaluating performance assessments. Simply put, in the process of implementing performance assessment, teachers, principals, and district office staff will have to change what they do and how they do it.

The papers selected for inclusion in this section represent a wide range of perspectives and suggestions from the change literature. Each offers a number of concepts that teachers and administrators should consider in planning for implementation and for facilitating the change process. Some of the papers were written by school practitioners, others by researchers who specialize in studying change. The papers range from those that report characteristics of the district, to those that discuss the role of the principal, to others that recommend strategies for monitoring implementation at the classroom level.

The papers divide into two major sets: (1) those addressing characteristics of schools and school districts that make a difference in implementation success; and (2) those describing concepts, models, and tools from change process research that have direct implications for achieving successful implementation of performance-based models. We have sequenced the papers within each set so that they flow from larger, more macro, perspectives, to the more micro and detailed. Each paper includes a list of references that are worth pursuing for further information. In addition, there are useful tools and strategies in the papers cited in the "Publications" portion of this guide's "Resources" section.[1]

## Characteristics of Districts and Schools That Make a Difference in Implementation Success

The characteristics of the school district—or macro—context within which school staffs will implement performance-based models can make a major difference. For some years there was an assumption that schools could bring about major improvements by themselves. This assumption has died hard, as more people are accepting that sustained success in school change must be linked to the larger system. The first three papers in this set address characteristics of the school district that make important differences in the degree to which schools and classrooms are able to implement innovative practices.

In the first paper, **Sam Stringfield** calls for systemic efforts to create schooling organizations (schools and districts) that can help all children achieve high, measurable goals. Such achievement, Stringfield believes, is possible in high-reliability systems that are characterized by 12 factors identified from Stringfield and colleagues' research in schools and review of the literature. The role of performance assessment, though not addressed per se in the paper, is obviously part of such systems.

The second paper, by **H. Dickson Corbett and Bruce L. Wilson**, delves into the activity of central office personnel. Although some want to debate the need for central office staff, this paper documents important elements of the role that central office personnel play in improvement at the school site. The paper is a report of research in five district central offices and concludes that developing and using instructional knowledge is key. Clearly, central office staff need to be involved, supportive, and consistently communicating the message that they expect and value instructional improvement. Including language about the importance of performance assessment in the instructional process, and indicating the importance of alignment among instruction, curriculum and performance assessments, will facilitate implementation of performance assessments.

The next paper, by **Sam Stringfield and Charles Teddlie**, shifts the focus to the school site. As part of their ongoing studies of effective schools practices in Louisiana, Stringfield and Teddlie have documented characteristics of schools and school districts that are related to successful change. They identify characteristics of school leaders and point out how the leaders act as important linkages between the schools and the district. As other researchers have reported, the authors have found that the principal makes a major difference in the degree to which teachers put new practices, including performance assessments, into place and sustain the practices in their classrooms.

The critical role of the principal is also the focus of the paper by **Terrence E. Deal and Kent D. Peterson**. As they point out, "Principals do more than lead change. They also manage its dynamics" (p. 1). The authors address two major roles of the principal: (1) attending to the nitty-gritty details, and (2) building a new culture and providing the vision. Being explicit about the importance of assessing what matters, and using high-quality performance assessments to do this as the leader goes about both aspects of this work, will facilitate the change process.

The last paper in the first set is by a nationally recognized principal, **Tim Westerberg** of Littleton High School in Littleton, Colorado. His high school received national attention for their systematic, long-term effort to develop and implement performance-based graduation requirements. He offers a number of thoughtful recommendations for others who are going in this direction. As he was writing this paper, the Littleton community elected three ultra-conservative candidates to the school board. The new board rejected the previously approved outcomes-based graduation requirements and reinstated the traditional course-hour requirements. Tim's important points and recommendations reflect wisdom gained through trial by fire. One important learning may be that the community must be an active partner as school systems develop and implement newer forms of assessment.

---

[1] Several of the papers in this section—Stringfield; Corbett and Wilson; Stringfield and Teddlie; and Deal and Petersen—have been edited down considerably from their original versions. This was done in order to delete lengthy case study descriptions with little relevance to student performance assessment, while retaining the generally applicable findings and conclusions regarding implementation and change.

## Concepts and Recommendations from Change Process Researchers

The second set of papers, written by scholars who study change processes in schools and other organizations, offer important concepts and, in some cases, models or theories for use in planning and facilitating implementation efforts. These papers offer research-based tools to use in thinking about the change process. School and district staff can apply the principles in nearly all settings: There are indeed some guidelines for bringing about change successfully, just as there are guidelines for developing performance assessments.

The first paper in the change process set, written by **Michael G. Fullan**, begins by emphasizing an ongoing theme: Change is in the hands of teachers. The paper then highlights some of the regularly observed barriers to change. The last half offers concise descriptions of numerous change strategies that can make a difference in the rate and degree of success. All of Fullan's advice is applicable to development and implementation of performance assessments.

In the paper that follows, **Shirley M. Hord** combines findings from change researchers and points out some of the key assumptions about change, continuing to focus on the importance of leadership. This paper concludes with a set of brief descriptions of change strategies that can make a difference.

The first "Issues...about Change" paper, by **Vicki Boyd**, returns to the now-familiar frame of "context." Whereas the first three papers in the first section dealt with characteristics of the district, Boyd's paper addresses the same ideas with more of a conceptual lens and a tie to the individual school. Be sure to check out the reference list on the last page.

The second "Issues...about Change" paper, written by **Shirley Hord**, reviews other important findings from research on school improvement processes. The findings and concepts presented in this paper move the discourse into the classroom. As researchers have pointed out for over 20 years (Hall and Hord 1987), a school cannot change until the teachers within it change. Again, the reference list on the last page of this paper includes many useful sources.

## Summary

The first seven sections of this resource guide address different aspects of *developing* performance assessments in the context of school restructuring. In this section we address issues of *implementation*. Having created the finest performance assessments in the context of school restructuring will make no difference unless teachers use them. The papers included here will be helpful in planning for change, facilitating implementation, and evaluating results. It is critical to remember that assessing outcomes without first documenting implementation is a sure way to achieve a conclusion of "no significant results."

In this guide we have made a distinction between the processes of *developing* performance assessment programs or systems and *implementing* those systems. While these processes are most often linear, in some settings they are cyclical or even simultaneous. In the first instance, a school or district examines, adopts, and then implements a program or system that another district has developed and tested. In the second case, a school or district explores an area of concern and then develops, tests, revises, and implements (with continuous revision) the program or process that particularly satisfies their need for change.

This latter schema seems to happen especially among a strong, collaboratively engaged faculty that is experienced in articulating and communicating their shared mission, in making broad-based decisions, and in sharing facilitation and support with each other. This kind of develop/implement/revise cycle is the focus of Southwest Educational Development Laboratory's research at an elementary school study site in Louisiana. Other states, such as Oregon in the Northwest Regional Educational Laboratory's region, where performance assessment is the target of widespread attention, also exemplify this approach.

In Oregon, teachers from around the state identified 11 desirable student outcomes, with teachers in different schools selecting different outcomes to address. A report from several of these teachers suggests the following sequence:

> 1. Focusing on a single outcome, we developed rubrics and task guidelines, using our past experiences with process-writing as a model.
> 2. We wrote tasks in all subject areas at all grade levels that will be used with students to demonstrate mastery of the outcome; several tasks were developed as alternatives.
> 3. We piloted the rubrics and tasks with teachers and students in our school. This field work required a school year for testing in each academic discipline and grade level.
> 4. We examined students' work carefully and revised the rubrics and tasks; staff at the state level also provided input for the revisions.
> 5. In the summer of 1994 we participated in weeklong institutes to share the documents to which we contributed and that included the work on all 11 outcomes.
> 6. The document was mailed to those districts whose personnel were not able to attend the institute. These notebooks were divided into the 11 outcome areas that included rubrics, task guidelines, examples of tasks, and examples of students' work.

We expect that some classroom teachers will access the notebook and use its items "as is." Others will use it as a starting point for developing tasks related to their own course content, demonstrating the more "messy" develop/implement/revise strategy. In either case, implementers will require the support and guidance of the school's leaders—the nature of which is fully articulated in the several papers in this section focusing on leaders.

Some of the papers in this section offer methods of addressing implementation from a district-level perspective. Several papers describe the importance of leadership at the school site *and* district level. Others offer strategies to assess implementation at the school and classroom levels. Each presents recommendations for making sure that implementation is successful. The reference lists at the end of each paper also warrant attention. They will be helpful in expanding the knowledge and resource base for implementation, which in the end is where differences will be made.

# VIII-2

# Attempting to Enhance Students' Learning: A Search for Valid Programs and Reliable, Systemic Implementation Supports

SAM STRINGFIELD
CENTER FOR RESEARCH ON THE EDUCATION OF STUDENTS PLACED AT RISK, JOHNS HOPKINS UNIVERSITY

If we know little else about school cultures, we know they tend to be stunningly stable and well-defended against virtually all forms of threats to "the way we were." (See, for example, Cuban 1993; Sarason 1990.) While the federal and state governments have called in ever-more-harsh tones for the restructuring of schools, those same institutions have flat-funded and often cut funds for research and evaluations of possible avenues to improving schools. As if that weren't enough, analyses of the changing American demographics indicate that during the 21st century nearly half of all school-aged U.S. children will live in conditions we now call "at risk" (Natriello, McDill, and Pallas 1990).

The challenge is deceptively simple: to get virtually all students to achieve at levels heretofore achieved by only a small percentage of our students. The proposals for meeting that challenge fall into three broad categories:

1. Raise academic standards and test to those new, high standards. This tells educators where to go, but not how to get there.

2. Use new programs. Such programs typically tell educators how to get there, but most offer precious little hard data to support their claims of efficacy.

3. Restructure schools. This outgrowth of (or, in some cases, reaction to) the school effectiveness movement focuses on site-based (e.g., school-level) change processes.

---

This paper has been edited down considerably from its original version. This was done in order to delete a lengthy case study description with little relevance to student performance assessment, while retaining generally applicable findings and conclusions regarding implementation and change. Copyright © 1995 Sam Stringfield.

Two striking features of all three responses are that (a) all decry the status quo and make considerable promises for their particular solution, and (b) few discuss the multi-level "how to" of making reform work.

A study of ten promising programs for improving the education of disadvantaged students, known as the Special Strategies Studies, was conducted to assess these programs' potential to improve the education of all students. A particular focus has been students who were not doing well in school, but all students at one grade in 25 separate schools have been followed for three full academic years.

The final report of this study is now in process at the U.S. Department of Education; however, first-year findings have been recently published (Stringfield et al. 1994) and are overviewed here:

1. *Virtually all of the programs studied appear to possess some clear and often unique advantages which might recommend them to some schools and school districts.* None of the programs seemed like particularly "bad" ideas, and all seemed to have some potential to address one or more problems that might be faced by local students and/or schools.

2. *A specific strategy was often chosen by administrators and/or teaching staff with little consideration of alternatives.* While there were examples within the study of district or school-level staff engaging in an extended needs assessment process and systematically searching for alternatives, there were more examples of a strategy simply being imposed from above, or of a school choosing a strategy after only word-of-mouth recommendation. In either case, there was rarely more than a cursory search for alternatives or third-party evaluations for the specific strategy being chosen.

3. *Strategies designed to affect students' whole school days typically resulted in higher levels of interactive instruction throughout the students' reading, language arts, and mathematics classes than programs which are adjunct to the regular day.* Adjunct programs often provided potentially valuable services to students during pullout periods, but typically did not result in high rates of student engagement during the remainder of the day.

4. *The "Special Strategy as received by students" varied within and across sites and programs in relation to each school's skill at implementing new programs, teacher expertise, and magnitude and intensity of intervention.* Whole school day observations of students often left very different impressions of the special strategy and its implementation at the school than would have been gleaned from a simple set of interviews of individual central administrators, principals, teachers, and students, or from reading published descriptions of the strategy.

5. *Externally developed programs requiring fidelity to a prespecified model required support from either model developers or other high-quality technical assistance.* The schools in which such strategies as Reading Recovery, CCC, Paideia, the Coalition of Essential Schools, Success for All, and the Comer School Development Program were well- implemented and were invariably the recipients of extended, often collaborative, training or assistance from the developers.

6. *District and state commitment to ongoing staff development was often key in facilitating special strategy implementation and institutionalization.* Regardless of the strategy type observed, local and often state support for ongoing professional development in that strategy consistently was associated with higher levels of implementation. Whether the program being implemented was seemingly self-contained (for example, pullout computer- assisted instruction) or a schoolwide effort at a philosophy-based approach (for example, the Coalition of Essential Schools), ongoing, long-term support for development of professional skills was critical to the health of the strategy-as-implemented.

7. *Active leadership was crucial to program implementation.* Implementing the various special strategies required somewhat different leadership skills. However, in every site visited by the Special Strategies teams, the principal's level of commitment to the chosen program, organizational skills, and ability to motivate self were relevant to the coherence of the program as implemented.

8. *Powerful contextual variables facilitated or impeded implementation of the various strategies.* All special strategies were subject to local variables that transcended particular strategies. Unanticipated school district budget distress was one of several issues that invariably produced negative impact on implementation efforts. Local access to strategy-specific technical assistance almost always had a positive impact.

9. *Specific implementation issues varied with the intended magnitude and scope of the special strategy on the structure of the school and the content of instruction.* While many of the findings relate to general issues in implementation, several issues relate to specific types of programs. The range of issues is illustrated in the following two examples. Reading Recovery requires a full year of intensive training for prospective first grade teachers before they are certified as Reading Recovery specialists, but the program does not include training for second grade teachers (although some schools offer this and more). Paideia schools require significant changes in the daily teaching patterns of all teachers, but these can be phased in over several years.

10. *With the possible exception of some adjunct programs, all special strategies visited continued to evolve at the various sites.* An easy mistake would be to imagine that any of the schools in the Special Strategies studies was "doing (its strategy) right." In almost every instance, schools continued to struggle with and evolve new methods for implementing their chosen program. These implementations were highly dynamic.

11. *Central to replication efforts would be the systematic exploration of the preconditions for implementation; the roles of such key staff as principals, faculty, and parents; the explicit relationship between the special strategy and instructional methods and curriculum; and the extra visible, as well as hidden, costs associated with implementation.*

## High-Reliability Organizations

What would high quality systemic support for school improvement look like? What are the things that parents, teachers, and principals could demand, and that whole systems could strive to supply?

In one sense, the America 2000 goals do not demand new programs. We already have curricula and instruction in place that allow many students to graduate from high school and to score high on international tests. We lack *reliability* of implementation. Not all students are getting remotely equal services.

Suppose we were to try to improve the U.S. schooling system by increasing the systemic reliability. What would this more reliable system look like? Research on high-reliability organizations (HROs) can be melded with the current generation of school effects research to provide at least partial answers. Diverse organizations required to operate with extraordinarily high reliability share the following characteristics:

1. *The most important shift in considering HROs is not in the specific characteristics of the organizations; it is the intellectual shift that precedes the evolution of the characteristics.* ***The most important characteristic of HROs is a perception, held by the public and the employees, that failures within the organization would be disastrous.*** Mid-air collisions of aircraft or extended power blackouts are invariably viewed as disastrous. Schools will increase the reliability of their work as the public and the professionals working within the schools come to view children's not learning to read as constituting individual travesties and public disasters. That situation is rapidly approaching in terms of public perceptions of schools. Dropouts and other ill-educated citizens are becoming disastrously expensive.

2. ***HROs require clarity regarding goals.*** *Staff in HROs have a strong sense of their primary mission.* Establishing and maintaining clear goals has been one of the most frequently cited characteristics of the school effects research base (Edmonds 1979; Good and Brophy 1986; Levine and Lezotte 1990). Stringfield and Teddlie (1988, 1991) found that a cacophony of "most important" goals often resulted in a lack of clear, unifying goals for the faculties and students in low-achieving schools. A focus on specific outcome measures clarifies objectives throughout an educational system.

3. ***HROs extend formal, logical decision analysis,*** *based on standard operating procedures: as far as extant knowledge allows.* At the time of publication of the first *Handbook of Research on Teaching* (Gage 1963), a sufficient body of educational research did not exist to rationally guide the development of standard operating procedures for schooling. In the absence of such knowledge, and in the absence of clear national or state goals, districts developed local procedures. Districts, local superintendents, and principals often abdicated these responsibilities to teachers, with virtually no monitoring of resulting practice. By the time the third *Handbook* (Wittrock 1986) was published, the research situation had changed. The rudiments of a science of education now exist. Practice has been slow to follow.

An important link between school effects research and HRO literature is provided by two studies of teachers' behavior in more and less-effective schools. Both Mortimore et al. (1988) and Teddlie and Stringfield (1993) found that teachers in high outlier schools behaved in ways more like those predicted by the teacher effects literature (Brophy and Good 1986). Perhaps as important, these studies found greater *consistency* among teachers in more effective schools. More teachers were more likely to be moving students through their lessons at a good pace, and fewer teachers allowed high rates of time off-task. Stringfield and Teddlie (1991) found that this higher rate of consistency on the part of teachers was clearly related to the behaviors of the principals, and not related to the income levels of the communities being served.

4. ***HROs recruit and train extensively.*** *This is true in part because at peak times, professional judgment is valued.* Kozol (1991) repeatedly observed that at-risk children need the highest quality teachers, yet often do not get them. There is little research on teacher, principal, and superintendent recruitment, and essentially none on the "effects" of recruitment. Research on the "how to" of effective training has made much greater progress (Showers, Joyce, and Bennett 1987; Louis and Miles 1990; Fullan 1991). Ward and Tikunoff (1989) and Kirby et al. (1991) have conducted detailed studies of teacher induction programs, both of which found school-level effects. However, much research on the long-term effects of

various induction and staff development programs has yet to be conducted.

Stringfield and Teddlie (1991) found that principals in positive outlier schools were more likely to take an intense interest in staff recruitment. They found that, by contrast, principals in low outlier schools passively accepted "what they (central administration) send us." This was particularly true in schools serving high numbers of at-risk students.

5. ***HROs have initiatives which identify flaws*** *in standard operating procedures and nominate and validate changes in those that prove inadequate.* The author is unaware of large-scale research on school or district-level efforts at systematic, organizational efforts to identify flaws within schools and correct them. However, Mortimore et al. (1988) and Stringfield and Teddlie (1991) described principal and staff actions in more effective schools that might serve to identify problems within schools and facilitate correction. Stringfield and Yoder (1992) present a case study of an exemplary school serving highly disadvantaged Hispanic children. That school consistently developed standard operating procedures and with equal frequency sought out methods for improving those procedures.

The Federal Aviation Administration carefully studies every airplane crash in the U.S. and uses those case studies to modify both technologies and human systems. In education, our tendency has been to shun dropouts. Yet, it is their experience that can inform efforts to improve reliability of schooling.

6. ***HROs are sensitive to the areas in which judgment-based, incremental strategies are required.*** *They, therefore, pay considerable attention to performance, evaluation, and analysis to improve the processes of the organizations.* U.S. education is currently undergoing increased emphasis on performance-based evaluation of teachers (e.g., Millman 1989).

Stringfield and Teddlie (1991) reported that principals in the more effective schools took teacher recruitment, development, and evaluation more seriously than did principals in the study's negative outlier schools. Most of the principals in the positive outlier schools had counseled-out, forced to transfer, or otherwise removed one or more teachers from their staff. That was rarely true of principals in negative outlier schools.

7. ***In high-reliability organizations, monitoring is mutual*** *(administrators and line staff)* ***without counterproductive loss of overall autonomy and confidence.*** In the U.S. there is currently minimal evaluation of teachers by administrators or fellow teachers, and virtually no evaluation of administrators by teachers. In recent school improvement studies, both Rumberger and Larson (1994) and Fine (1994) present systems in which peers become much more aware of each other's teaching. This is a step toward institutional reliability.

8. ***HROs are alert to surprises or lapses.*** *The experience of HROs is that small failures could cascade into major system failures, and are hence monitored carefully.* The simple freezing of a rubber O-ring caused a series of malfunctions which led to the Challenger space shuttle disaster. A squirrel working his way into an electric box "caused" the great New York and East Coast power failure. In both cases, it was the fact that the initial, simple failure was able to have *cascading effects* that turned a small problem into a disaster. It is this cascading failure, repeating daily throughout our education system, that has created the potential economic disaster that would result from the continuation of downward income trends noted by Murnane and Levy (1993).

The author is unaware of studies focused specifically on schools' responses to surprises or lapses. It is easy to imagine that Mortimore et al.'s (1988) *"purposeful leadership,"* "maximum communication between teachers and students," and "record keeping"—all found to be characteristics of more effective schools—would lead to quicker response to surprises or lapses.

Similarly, Stringfield and Teddlie's (1991) "attention to daily academic functioning" as a school-level predictor of school effects, might lead to quicker discovery of and response to lapses.

9. ***HROs are hierarchically structured, but during times of peak loads, HROs emphasize a second layer of behavior which emphasizes collegial decision making regardless of rank.*** *This second mode is characterized by cooperation and coordination. At times of peak activity, line staff are expected to exercise considerable discretion. Especially during times of peak performance,* ***staff are able to assume a close interdependence.*** *Relationships are complex, coupled, and sometimes urgent.*

U.S. schools are hierarchically structured. Expectations during "peak" times have not been systematically studied but probably vary greatly. During busy times in several Special Strategies schools, the "basic" components of the school (e.g., regular classroom teachers) were protected at the expense of all "special" programs (e.g., counselors were called out to staff the office if the principal had to be away). My experience is that during peak times at some schools, accommodations are made, everyone pitches in, and important functions continue to operate.

Getting rid of hierarchy in education may not be an important goal. Allowing exceptions to bureaucratic rules and communication patterns at moments when side-stepping the hierarchy can save a student from illiteracy may be what is needed.

10. ***Equipment is maintained in the highest working order.*** *Responsibility for checking the readiness of key equipment is shared equally by all who come in contact with it.* This is not true in most U.S. schools, where nonfunctioning equipment often sits unrepaired for months or years. A striking feature of equipment maintenance in many schools is the extent to which unnamed persons beyond the school are often viewed as being responsible for maintenance.

11. ***HROs are invariably valued by their supervising organizations.*** There is some evidence (e.g., Wimpelberg, Teddlie, and Stringfield 1989) that school districts provide more attention and support to some schools than others, and schools in the least advantaged neighborhoods often receive the least attention and local support. Stringfield and Teddlie (1988) have observed that some school districts, rather than retraining or terminating manifestly unskilled teachers and principals, assign them to schools serving the least advantaged families in the districts. Such actions do not indicate high valuing of *all* students and schools by central offices.

12. ***Short-term efficiency takes a back seat to very high reliability.*** U.S. education has spent much of the last 30 years attempting to become more efficient, and much public dialogue concerns ridding education of "wasteful management."

Reading Recovery is an interesting example of a program which places much more weight on reaching long-term goals than on short-term efficiency. Reading Recovery is not efficient, but several studies indicate it is effective.

High-reliability organization literature describes 12 characteristics of organizations which are required to operate trials without cascading errors. This is the requirement increasingly being made of schools. Several school effects studies and the Special Strategies Studies suggest linkages with the HRO literature. However, for a typical school serving large numbers of at-risk students to become an HRO, it would need to make major changes. Some of those changes would require support from above the school.

Schools and administrators would need to clarify goals; create and clarify standard operating procedures regarding curriculum, instruction, and the handling of various student and task management issues; greatly increase goal-directed staff development; further open access to rule-making; improve two-way staff evaluation procedures and practices; be clearer and more flexible in dealing with situations in which special needs of children must be met (bureaucratic concerns must occasionally be subsumed under practical, extraordinary needs); replace teachers' isolation with more collegial working relationships; and finally, a school would need to allocate funds not just to procure new equipment, but to maintain existing equipment. These will be difficult, but necessary, steps.[1]

## REFERENCES

Brophy, J.E., and T. Good. (1986). "Teacher Behavior and Student Achievement." In *Handbook of Research on Teaching*, 3rd ed., edited by M. Wittrock. New York: Macmillan.

Cuban, L. (1993). *How Teachers Taught*. 2nd ed. New York: Teachers College.

Edmonds, R. (October 1979). "Effective Schools for the Urban Poor." *Educational Leadership* 37, 1: 15-27.

Fine, M. (1994). "Chart[er]ing Urban School Reform: Philadelphia Style." In *Educational Reforms for At-Risk Students*, edited by R. Rossi. New York: Teachers College Press.

Fullan, M. (1991). *The New Meaning of Educational Change*. New York: Teachers College.

Gage, N. (1963). *Handbook of Research on Teaching*. Chicago: Rand McNally.

Good, T. and J. Brophy. (1986). "School Effects." In *Handbook of Research on Teaching*, 3rd ed., edited by M. Wittrock. New York: Macmillan.

Kirby, P., S. Stringfield, C. Teddlie, and C. Wimpelberg. (1991). "Teacher Induction into More and Less Effective Schools." Paper presented at the meeting of the International Congress for School Effectiveness and Improvement, Cardiff, Wales.

Kozol, J. (1991). *Savage Inequalities: Children in America's Schools*. New York: Crown.

Levine, D., and L. Lezotte. (1990). *Unusually Effective Schools*. Madison, Wisc.: National Center for Effective Schools R & D.

Louis, K., and M. Miles. (1990). *Improving the Urban High School: What Works and Why*. New York: Teachers College Press.

Millman, J. (1989). *Handbook of Teacher Evaluation*. 2nd ed. Beverly Hills, Calif.: Sage.

Mortimore, P., P. Sammons, L. Stoll, D. Lewis, and R. Ecob. (1988). *School Matters*. Somerset, England: Open Books.

Murnane, R., and F. Levy. (1993). "Why Today's High School Educated Males Earn Less Than Their Fathers Did: The Problem and an Assessment of Responses." *Harvard Educational Review* 63, 1: 1-19.

Natriello, G., E. McDill, and A. Pallas. (1990). *Schooling Disadvantaged Children: Racing Against Catastrophe*. New York: Teachers College Press.

Rumberger, R., and K. Larson. (1994). *Keeping High-Risk Chicano Students in School: Lessons from a Los Angeles Middle School Dropout Prevention Program*. New York: Teachers College Press.

Sarason, S. (1990). *The Predictable Failure of Educational Reform*. San Francisco: Jossey-Bass.

Showers, B., B. Joyce, and B. Bennett. (1987). "Synthesis of Research on Staff Development: A Framework for Future Study and a State-of-the-Art Analysis." *Educational Leadership*, 45, 3: 77-87.

Stringfield, S., and C. Teddlie. (1988). "A Time to Summarize: Six Years and Three Phases of the Louisiana School Effectiveness Study." *Educational Leadership* 46, 2: 43-49.

---

[1] For further discussion of high reliability within education contexts, see Stringfield et al. 1994 and Stringfield and Herman 1995.

Stringfield, S., and C. Teddlie. (1991). "Observers as Predictors of Schools' Multi-Year Outlier Status." *Elementary School Journal* 91, 4: 357-376.

Stringfield, S., L. Winfield, M. Millsap, M. Puma, B. Gamse, and B. Randall. (1994). *Urban and Suburban/Rural Special Strategies for Educating Disadvantaged Children: First Year Report.* Washington, D.C.: U.S. Department of Education.

Stringfield, S., and N. Yoder. (1992). "Toward a Model of Elementary Grades Chapter 1 Effectiveness." In *Students at Risk in At-Risk Schools*, edited by H. Waxen, J. Deflect, J. Anderson, and P. Baptist. Newbury Park, Calif.: Sage.

Teddlie, C., and S. Stringfield. (1993). *School Effects: The Roles of Principals, Teachers, and Students Over Time.* New York: Teachers College Press.

Ward, B., and W. Tikunoff. (1989). *New Teacher Retention Project.* Los Alamitos, Calif.: Southwest Regional Educational Laboratory.

Wimpelberg, R., C. Teddlie, and S. Stringfield. (1989). "Sensitivity to Context: The Past and Future of Effective Schools Research." *Educational Administration Quarterly* 25, 1: 82-107.

Wittrock, M., ed. (1986). *Handbook of Research on Teaching.* 3rd ed. New York: Macmillan.

# VIII-3

## The Central Office Role in Instructional Improvement

H. DICKSON CORBETT AND BRUCE WILSON
RESEARCH FOR BETTER SCHOOLS

The quality of education a school district offers is the extent to which *all* of its students become successful learners, with successful learners being knowledgeable, empathetic, self-determined, and strategic (Jones and Fennimore 1990). If students' opportunities to engage in activities that encourage them to become *knowledgeable, empathetic, self-determined,* and *strategic* qualitatively vary among classrooms, programs, and/or schools, then students are at the mercy of the district's assignment process as to their chances of succeeding. We argue in this paper (1) that the instructional improvement role of the school district central office is to insure that every building and classroom substantially promotes students' chances of becoming successful learners, and (2) that, based on our research in five school systems, a key way of accomplishing this is to facilitate staff members' access to and use of important bodies of instructional knowledge.

### Perspective

In most of the school systems with which we have experience, there is a hit or miss character to the quality of education. Depending on which school a child attends and to which classrooms the child is assigned, the student will encounter a varied array of programs and activities. The consequence is that all students may develop some of the characteristics of successful learners, but only some students will be fortunate enough to follow a route through the system that encourages the development of all four. From the students' perspective, then, the quality of their educational experiences rests on the "luck of the draw." The central office's instructional role is to remove this luck factor from the instructional program; that is, to insure that idiosyncratic variations in programs, people, and policies do not result in systematic differences in the quality of education for children. The central office must be able to assure itself and others, for example, that (1) students with

---

Adapted from: H.D. Corbett and B.L. Wilson, "The Central Office Role in Instructional Improvement," *School Effectiveness and School Improvement* 3, 1 (January 1992): 45-68. Reprinted with permission. This paper is shorter than the original version because the authors have deleted a lengthy case study having little relevance to student performance assessment while retaining generally applicable findings and conclusions about implementation and change.

minimal basic skills are not relegated to a minimalist education; (2) "average" students have the same richness of education as the gifted; (3) students with a first-year teacher will have no less, or no more, access to engaging activities than students with a veteran teacher; and (4) a student with one fourth grade teacher will neither be advantaged nor disadvantaged in comparison to a student with another fourth grade teacher.

To achieve such consistency, the central office has to make certain that every instruction-related decision is informed by knowledge about how best to encourage successful learning. The "luck of the draw" becomes a major determinant of students' opportunities to succeed where there are vast variations in knowledge and skill across practitioners in the system. On the other hand, assignment to buildings and classrooms becomes a minor determinant of success where variation in practitioners' knowledge *and* skill is minimal and where practitioners' access to new knowledge and practitioners' opportunities to improve their skill are uniformly high.

A central office may weave important instructional knowledge into practice through a variety of mechanisms: hiring the most knowledgeable and skilled people possible, promoting staff development, concentrating on curriculum development *and* implementation, improving supervision of staff, conducting program evaluations, and/or relying on wishful thinking. Whatever combination of mechanisms a central office uses, its core concern is whether good instructional information is accessible and used in day-to-day school life. Our work examines the means through which central offices promote uniform access and use to instructional knowledge that purports to increase practitioners' ability to reach all children.

A word of caution: Having a uniform distribution of knowledge and skills does not preclude individuals from using a variety of personal styles, classroom activities, *and/or* content topics with students. Uniformity here refers only to the distribution and use of a knowledge base that can inform practitioners' work. A sameness in the knowledge base practitioners draw upon does not imply a sameness in the actions that flow from the knowledge base; rather it helps to establish a single set of criteria by which decisions are made. So, this perspective does not constitute a case for standardization and/or centralized decision making.

## Study Purpose

The primary purpose of the study was to identify the ways in which central offices encouraged administrators and teachers to access and use important instructional knowledge. Specifically, we examined how five central offices (1) organized instructional improvement efforts, (2) led those efforts, (3) developed normative support for improvement, and (4) affected the extent of teachers' and administrators' access to and use of instructional knowledge attributable to those activities.

For definitional purposes, "instructional improvement organizations" means the regular patterns of rules, roles, and relationships pertaining to planning, implementing, and assessing activities intended to improve instructional practice. That is, we attended to the formal and informal vehicles through which staff members routinely acquired knowledge and to the structural mechanisms (resources, policy, incentives, schedules, communication networks, etc.) through which use of this knowledge was encouraged. "Central office leadership behavior" means the ways in which central office staff members influenced the instructional behavior and beliefs of practitioners. Our concern was with the specific actions of central office administrators. "Improvement norms" means the shared understandings among staff members about how important improvement is to the system and about how it should be conducted. We were most interested in the extent to which expectations developed for how a school system, rather than individuals, should go about improving.

We use information about a district's organizational arrangements, leadership behavior, and shared improvement norms to explain the extent to which teachers and administrators had access to and used instructional knowledge. Instructional knowledge refers to sets of instructional theories that have been generally accepted within the educational community as valid ways of promoting successful learning for all students.

We define "access" as the extent to which staff members actually availed themselves of opportunities to acquire knowledge. It was not enough for the district to generate improvement opportunities that theoretically were open to everyone; for the information contained therein to be of value to the district, people actually had to come into contact with that information. Thus, access meant that staff members were physically present when a significant portion of the knowledge base actually was conveyed.

An extensive literature has grown up around the complexity of determining how individuals use knowledge to construct action. An important lesson contained in this literature is that knowledge use can take varied forms (e.g., Hall & Loucks 1977; Weiss 1977). We categorize educators' "use" in three ways: (1) acknowledged incorporation of instructional knowledge into instruction-related responsibilities, (2) increased awareness of ideas or practices that may become translated into practice in the future, and (3) perceived reinforcement of practices/perspectives already present in a person's professional repertoire. We also examined whether educators believed a common language for discussing instruction was emerging in their districts—an indicator of a "shared" aspect to use that went beyond the individual classroom.

## Discussion of the Case Summaries

The case summaries we compiled based on our research highlight differences in five school districts' approaches to school improvement: controlled choice, top-started/bottom-run, evolutionary change, bottom-started/bottom-run, and countering diversity. At the same time we have noted variations in educators' access and use of instructional knowledge. This section summarizes these variations and then posits some explanations for them, based on differences in the school systems' organization for improvement, central office leadership behavior, and improvement norms.

### Differences in Access and Use

The following table (Figure 1) compares the districts on five categories: veteran teachers' access, newly hired teachers' access, administrators' access, teachers' use, and administrators' use. We assigned an adjective to each district for each category based on our aggregate assessment across all interviews. With respect to access, the terms "all," "most," and "few" refer to our estimates of the number of staff members relative to the totality of people in their respective role group who actually availed themselves of opportunities to encounter new knowledge. It does not refer simply to whether there were opportunities available. All of the districts made opportunities available to all of their staff members; the difference was in how many staff members actually participated in such activities beyond traditional inservice days. Concerning use, "widespread" means that many staff members across all levels and job specialties incorporated the information into their instruction-related responsibilities; "spotty" means that either a smattering of staff members districtwide incorporated the information into instruction-related responsibilities or there were pockets (a school, a program, a grade level, or a department) where use was widespread.

In District 1, every teacher had access to knowledge about instructional techniques through the staff development and new-teacher induction programs. Likewise, every administrator had access through initial administrative training and serving as trainers. Based on our interviews, many teachers from a variety of grade levels and program areas had discovered ways to incorporate the information into their instructional activities, and all administrators used the information as the basis for comments to teachers following classroom observations. Consequently, a common language to discuss instruction had developed in the system.

Teacher access in District 2 was similar, due to the initially required participation in Madeline Hunter training and to the teacher induction program. Most of the veteran teachers also seemed to have participated in at least one of the voluntary programs. All of the administrators took part in the Hunter training; however, only some were active in some of the current initiatives. Teacher use was found across all grades and departments. Actual administrative use was most evident at the building level in teacher evaluations and faculty meetings. Central office administrators were in less of a position to demonstrate use of new knowledge in instructional settings because of the way they had come to define

Figure 1

PATTERNS OF ACCESS TO AND USE OF INFORMATION ABOUT EFFECTIVE INSTRUCTION

| District | Current Teachers' Access* | New Teachers' Access | Administrative Access | Teachers' Use | Administrative Use | Common Language |
|---|---|---|---|---|---|---|
| 1 | All | All | All | Widespread | Widespread | Yes |
| 2 | All | All | All | Widespread | Widespread<br>Bldg. level | Yes |
| 3 | Most | Few | Some | Spotty | Spotty | No |
| 4 | Some | All | All | Spotty | Widespread<br>Bldg. level | Multiple |
| 5 | Some | All | Most | Spotty | Spotty | Beginning |

*Districts are ordered on the basis of current teachers' access.

BEST COPY AVAILABLE

their roles as initiators and supporters of improvement. Nevertheless, the development of a common language of instruction seemed to be taking place in this district.

We found that an increasing number of District 3 teachers were beginning to attend outside conferences and workshops. The same opportunities were available to newly hired teachers, but the absence of an instructionally substantive induction program and the proclivity of new teachers to "stay close" to their students suppressed these teachers' participation. All administrators had access to considerable information relevant to the management of their schools: it was less clear that all administrators had participated in activities with direct instructional implications. With the exception of teams at the middle school and one elementary school, the "spotty," rating meant that use of certain information was almost randomly distributed across the district. In the two schools where idea sharing was reinforced both formally and informally, pockets of use emerged. There was no mention of the emergence of a common language.

In District 4, veteran teachers participated in knowledge acquisition activities as a group less than new teachers as a group—an anomaly attributable to voluntary workshops and a required new-teacher induction program. All administrators, plus the department chairpersons, attended monthly instructional leadership sessions. Thus, their access was high. Use among teachers was spotty, given that only some availed themselves of staff development opportunities. Administrative use was widespread at the building level as indicated by incorporating "hunter-esque" issues into the evaluation form and occasionally devoting faculty meetings to instruction. Central office use was less, an expected consequence of the superintendent's acknowledged distance from day-to-day school activities. It was most appropriate to say that several instructional languages were sprouting within specific pockets of use.

District 5's curriculum director hoped that every teacher would participate in the system's voluntary summer workshop. To date, only about half had. New teachers received a truncated version of the workshop. Administrators were encouraged to attend various activities both within and outside the district; our information was that most had attended at least one session of the summer instructional effectiveness workshop. Use of new instructional knowledge gained through staff development appeared to be substantial *among those staff members who participated,* but relative to the district at large, such usage had a "spotty" appearance, with some buildings having a much greater proportion of teachers' evidencing use than others. The overall effect was to create pockets of use within the system; common ways to talk about instruction were cropping up in these niches.

### Promoting Access and Use Through a Consistent Message

Looking at the five approaches as a whole, "controlled choice" (District 1) and "top-started/bottom-run" (District 2) were associated with the most extensive levels of access and use among teachers and administrators and in both situations a shared instructional language sprouted. The distinguishing characteristic of the two approaches was the presence of a consistent message that instruction and its improvement were foremost in the local educational enterprise. All five districts obviously wanted staff members to take instruction seriously, but the *consistency* with which that message was communicated as *the* top priority varied. This consistency depended on (1) the degree of administrative involvement in instructional improvement initiatives; (2) the allocation of resources—especially staff time—to improvement purposes; and (3) the types of communication opportunities staff members had to discuss and demonstrate improved instruction. In Districts 1 and 2, involvement, support, and communication meshed to alert staff members that instructional improvement was unavoidable. Each of the other districts had distractions that weakened interview subjects' sense of an unquestioned commitment to instruction—for example, the remoteness of the central office from the classroom and almost exclusive reliance on volunteers. The difference was between a message that said "people will improve in this district" and one that said "people may improve in this district, if they choose to do so."

### Administrator Involvement

When administrators rolled up their sleeves and jumped with both feet into the improvement arena, teachers noticed. Such involvement took several forms: actually participating in professional development as a learner, daring to model effective instructional techniques in front of other adults, and appropriately applying instructional knowledge in classroom observations. This factor also seemed to be the catalyst that moved use from being an individual to a shared phenomenon, in terms of developing a vibrant common language for instruction. (It is interesting to note that in both Districts 4 and 5 an indication of a common language showed up in buildings where principals behaved much as all District 1 administrators did.) Teachers began to view principals and central office people as knowledgeable about instruction. As Dornbusch and Scott (1975) demonstrated, the more knowledgeable an evaluator is perceived to be by the person being evaluated, the more likely that the evaluation will be perceived as valid and a basis for altering behavior. Thus, while some administrators became involved in some professional development activities in each of the five districts, the difference in Districts 1 and 2 was the extent and intensity of involvement. All administrators were *engrossed* in improvement in District 1. Fewer administrators were directly engrossed in improvement in District 2 but the system had structured opportunities for teacher leaders to assume some of the modeling and feedback roles central office administrators played in District 1.

### Use of Resources

Resources included money spent on instructional improvement, material support given to these initiatives, and the judicious allocation of time for instructional improvement. All five central offices were viewed as being highly supportive of curriculum and professional development. But the type of resource use that seemed to be most directly related to differences in staff access and use was how the districts dealt with staff time to learn about effective instruction, that is, whether substantial staff development occurred *during the school day*. While all districts occasionally employed this strategy for a handful of teachers, Districts 1 and 2 trained all or most of the staff in this manner over the course of several years. In the other three districts there were very dedicated groups of teachers and administrators who pursued new instructional ideas either after school, on weekends, or during the summer. They were very enthusiastic, but generally voiced frustration in getting the knowledge to spread across the full faculty.

### Communication

Our interviews were liberally dotted with comments about the value of just hearing others discuss their work. "It lets you know you're not alone," "You realize that what you're doing is okay," and "You can always pick up a few tidbits to take back to your class" were representative statements. Wrapping these cherished moments within the framework of a specific instructional topic became a powerful means of engendering greater access and use of the information, and the power of communication opportunities was augmented by having a mix of formal and informal forums. The formal situations insured that discussion could occur; the informal enabled staff members to probe application issues that either they did not have time for in formal settings or they were uneasy about bringing up "in public." The combination of both let people know, almost every time they turned around, that improvement was both expected and "okay." The combination also boosted the number of occasions when staff members could try out and agree on common terminology for labeling classroom actions and consequences.

Both Districts 1 and 2 supported formal and informal communication vehicles that touched on all staff members. District 1 enabled both to occur mostly between teachers and administrators, through the required staff development activities, curriculum review committees, and the high number of informal central office observations of teachers. District 2, on the other hand, balanced the formal situations involving both administrators and teachers with a lot of chances for teachers to talk with one another, either as team members or peer coaches. Districts 3, 4, and 5 also had a mix of opportunities, but rarely did the same group of teachers have both. That is, not all teachers that attended formal meetings had the chance to talk in their buildings with others who were interested in, and/or had accessed, that same information. Similarly, teachers with considerable opportunities to talk informally with one another may or may not have availed themselves of professional development activities.

## Conclusion

Our assumption is that the extent to which important bodies of instructional knowledge permeate schooling makes a substantial difference in whether school districts ultimately are successful in enabling all students to become successful learners. Thus, improving educators' access to and use of this information is an intermediate—but critical—step on the way to reducing the "luck of the draw" in assigning students to buildings and classrooms as a major influence on which students become successful, because more students will likely encounter classroom situations in which the understandings and skills needed to engage them and enable them to succeed are present. We conclude (1) that the central office's role in instructional improvement is most effectively enacted through giving a consistent and clear message that instruction and its improvement is the primary expectation for adults in the system, and (2) that the consistency of this message is a function of involvement, support, and communication.

Involvement, support, and communication are probably overworked descriptors of what instructionally effective central offices do. Yet, the educators we studied pointed to the instructional improvement salience of these factors time and time again. But the educators warned sternly that it was not enough for some administrators to occasionally be involved; the power of involvement drew strength from the multifaceted participation of all administrators. Educators cautioned that general central office support for diverse and after-school professional development activities, while welcomed, fell short as a promoter of districtwide improvement; it took during-the-day, focused activities to channel enthusiasm and effort into a common direction. Educators appreciated an occasional chance to talk with one another, but they demonstrated convincingly to us that it took frequent formal and informal discussion opportunities for more than cursory use of new ideas to occur. Finally, educators were unequivocal in saying that widespread administrator involvement, focused allocation of resources, and multiple communication vehicles all had to be combined to generate a consistent message that instructional improvement was the local priority.

## REFERENCES

Dornbusch, S., and R. Scott. (1975). *Evaluation and the Exercise of Authority*. San Francisco: Jossey-Bass.

Hall, G., and S. Loucks. (1977). "A Developmental Model for Determining Whether the Treatment Is Actually Implemented." *American Educational Research Journal* 14, 3: 263-276.

Jones, B., and T. Fennimore. (1990). "The New Definition of Learning." In *Restructuring to Promote Learning in America's Schools*. Elmhurst, Ill.: North Central Regional Educational Laboratory.

Weiss, C. (1977). "Research for Policy's Sake: The Enlightenment Function of Social Research." *Policy Analysis* 3: 531-545.

# VIII-4

# A Time to Summarize: The Louisiana School Effectiveness Study

Sam Stringfield
Center for Research on the Education of Students Placed at Risk, Johns Hopkins University

Charles Teddlie
Louisiana State University, Baton Rouge

## How Schools Become Effective

In schools that eventually become effective, an instructional leader or leadership group, ideally though not necessarily the principal, arrives (or emerges). This person or group has a vision of what the school and its students could become, involving higher levels of learning and more humanistic processes than present realities would seem to justify. Though the leaders may not consciously know it in the beginning, they are prepared to work very long hours for years to achieve their vision.

Very early after arrival, the principal or group conducts an instructional audit of the school, either alone or with the staff. This set of steps resembles the following questions:

1. Is the school a safe place for children and staff?
2. Is the school a healthy and orderly place?
3. Is the school achieving "the basics"?
4. What is the status of the textbook series and supplemental materials within and across grades? Teaching and homework? Curriculum and instruction?[1]
5. Is the school providing additional enriching and stimulating activities for its students?
6. Is the school involving parents and the larger community in ways that are good for the students and for the community itself?

To make an intelligent choice of a beginning point for attaining the vision, the principals and the teachers must collect accurate knowledge of their schools' individual positions within the audit sequence. Then they know where to begin setting goals and marshaling resources to attain their goals.

---

This paper has been adapted from S. Stringfield and C. Teddlie, "A Time to Summarize: The Louisiana School Effectiveness Study." *Educational Leadership* 46, 2 (October 1988): 43-49. This version is considerably shorter than the original (see also Teddlie and Stringfield 1993); we have retained generally applicable findings and conclusions regarding implementation and change, but deleted a lengthy case study not relevant to performance assessment.

Although their awareness of research on teacher effectiveness varied from moderate to nonexistent, all principals in the Louisiana study exercised the commonsense notion that hard work leads to success. Regardless of the words used to describe the desired state, all expected to see students on task when they visited classrooms. And they were active in teacher selection, classroom observation, employment, and dismissal. From the beginning, these leaders chose new teachers and aides with great care. They looked for "spark" or "energy" and were often unconcerned with years of teaching experience or advanced degrees.

Wherever resources were available, effective principals became increasingly active in targeting staff development for some, occasionally all, of their teachers. This targeting was largely based on interests and needs they identified from their frequent in-class observations and subsequent conferences with teachers.

In addition, teachers who did not meet reasonable instructional standards were put on probation and provided assistance; they then either exhibited improvement, or were transferred to another school or district, or were fired.

The leaders emphasized instruction in other ways. For example, most effective principals had developed a minimum daily homework expectation. Also, programs such as Chapter 1 and special education were thoughtfully coordinated with the regular program. The teachers became aware of each other's practical specialties and viewed each other as resources on those topics. Teachers were often encouraged and sometimes required to visit each other's classes.

An interesting attribute of these principals is that they were never satisfied. They wanted more for their students, and they were continually looking for it. Growth for its own sake seemed to become one of the school's overall goals.

## How Schools Become Ineffective

By contrast, the ideal principal for the development of an *ineffective* school may very well mouth platitudes about child-centered education but actually became a principal in order to get out of the classroom. Occasionally this person had been a principal at another, probably a middle-SES school. If that were the case, the situation had become sufficiently unappealing to the staff or the parents that the superintendent transferred the principal to a different school, often a low-SES status school, where the parents may not complain as much.

This kind of principal envisions the job responsibilities as bureaucratic, more or less going through the motions of administration without commitment to any sort of vision. This principal may speak at length about the quality of the teaching staff, but in fact has not regularly visited the classes and actually judges teachers in terms of how little trouble emanates from their classrooms. Such a principal takes little interest in curriculum and avoids activities that have the potential to "stir up trouble." This passive stance usually rules out instructional leadership.

This type of principal accepts whatever prospective teachers are sent to him or her, often not even interviewing the new hires before they arrive for work. When interviewed, these principals reported having little or no voice in staff hiring. In our study the greatest variance in principals' self-reported perceptions of control over hiring was not between districts but within. That is, two principals from a single district would often give dramatically differing accounts of the degree of latitude the district allowed them in hiring. This perception was an excellent predictor of school effectiveness, especially in low SES-schools.

In addition, principals at ineffective schools almost never fired a staff member or even forced the transfer of one, despite the annual "floating" of ineffective teachers from school to school, called the "dance of the lemons" by Bridges (1986). Since highly effective principals do not accept their share of these teachers, a disproportionate share of them eventually come to work at an ineffective principal's school. Ineffective teachers usually like working for the ineffective principals, because they are left alone in their rooms to do as they please. Working with increasingly ineffective colleagues proves discouraging to the more effective teachers, who in time request transfers or drop out of teaching.

In time the school develops a reputation as an unpleasant place in which to work. Such a reputation can develop quite rapidly if the building serves an economically disadvantaged neighborhood. At that point, competent teachers within the system will actively resist transferring to the school.

---

[1] If, from the perspective of a below average student, the lessons over a pattern of years don't follow logically (e.g., different teachers use different texts), then a move is begun to standardize the text series and supplement them with additional materials. This scrutiny of materials provides one entry point for staff development. Several principals reported that most teachers exhibit no resistance when receiving inservice instruction on how to teach a new text and curriculum; yet the same teachers will take exception to someone trying to teach them the same principles about the present curriculum. It is, they observed, less threatening and more productive to work toward changes in instruction at the moment of changes in curriculum.

In a low-SES context, the staff develops an elaborate set of rationalizations for their behavior and for the school's performance, such as "Nobody could teach these kids" and "With parents like these, what do you expect?" These excuses serve to justify consistently poor performance. The principal reinforces these norms by writing "good" evaluation reports on all the teachers.

In an upper middle class ineffective school, teachers and principals observe that their students are scoring above the national average and congratulate themselves on a job well done. They do not compare performance to potential.

In all but the most productive districts, the presence of ineffective schools serves a valuable function. With these schools tacitly accepted as "dumping grounds," the central office staff are spared the arduous effort and the risks of dealing with incompetent teachers and principals, their unions, lawyers, and the courts.

Finally, the process of remaining ineffective is relatively resistant to school improvement efforts. In the absence of a strong internal support system, including norms for excellent performance and hard work and an emphasis on growth, any externally imposed improvement project will eventually disappear.

## REFERENCES

Bridges, E. (1986). *The Incompetent Teacher.* Philadelphia: Falmer.

Stringfield, S. (1994). "Identifying and Addressing Organizational Barriers to Reform." *Schools and Students At-Risk*, edited by R. Rossi. New York: Teachers College Press.

Stringfield, S. and R. Herman. (January 1995). "The Barclay/Calvert Experiment: At-Risk Students Get a Highly Reliable Implementation of a Proven Program, and Respond with Dramatically Higher Achievement." Paper presented at the meeting of the International Congress for School Effectiveness and Improvement, Leeuwarden, The Netherlands.

# VIII-5

# The Principal's Role in Change: Technical and Symbolic Aspects of School Improvement

TERRENCE E. DEAL
PEABODY COLLEGE OF VANDERBILT UNIVERSITY

KENT D. PETERSON
UNIVERSITY OF WISCONSIN-MADISON

## Policies or Purposes: The Principal's Role in School Improvement

People are demanding more and more from their public schools. They want improvements that serve children, especially those from poor or minority backgrounds. These improvements will often require major changes in existing patterns and practices. The changes will succeed only if school principals, and others, provide the needed leadership.

In this paper we delineate the important leadership roles that principals play in school improvement.

Principals do more than lead change. They also manage its dynamics. Their roles in school improvement have strong managerial components as well as powerful leadership imperatives (for a similar argument, see Wilson and Firestone 1987). On the other hand, they are symbolic leaders, building a new culture and providing the vision and passion needed for improvements to occur. If they are too rational, nothing will change; if too spiritual, nothing may happen.

---

Adapted from T.E. Deal and K.D. Peterson, "The Principal's Role in Change: Technical and Symbolic Aspects of School Improvement." Madison, Wisc.: University of Wisconsin-Madison, Wisconsin Center for Education Research, National Center for Effective Schools, 1993. Reprinted with permission. This paper has been edited down considerably from its original version in order to delete a lengthy case study description with little relevance to student performance assessment, while retaining generally applicable findings and conclusions regarding implementation and change.

## Blending the Two Perspectives

Management and leadership are not antagonistic; they are interconnected and symbiotic. Policies or practices can also be poetry and ritual. Improvement activities can accomplish measurable objectives as well as signal important values. Transforming a decrepit curriculum into a smoothly flowing one reinforces the value of quality teaching.

In our view, successful principals are both good managers and good leaders. Either that or they find ways to make sure that others compensate for what they do less well. Day in and day out principals have to attend to administrative details from budget reports to bus duty. They also have to find symphonic harmony in the confusion of values that exist in most school communities. Highly effective principals know that they can do at least two things with one action. In attending to administrative details, they send important symbolic messages of what is important. In working on managerial aspects of school improvement, they communicate important values and beliefs. By attending a graduation ceremony or telling a story, one can simultaneously accelerate the pace of change and accentuate learning outcomes.

## Technical Management

There are a number of technical activities that foster change and improvement (see Figure 1 below). They are complex and demanding.

FIGURE 1

TECHNICAL ACTIVITIES TO FOSTER SCHOOL IMPROVEMENT

| | |
|---|---|
| 1 | Developing a clear goal |
| 2 | Collecting and analyzing school outcome data |
| 3 | Designing an efficient planning model |
| 4 | Developing a plan to achieve school goals |
| 5 | Selecting strategies and processes for implementing the plan |
| 6 | Gathering necessary resources |
| 7 | Providing training and staff development |
| 8 | Coaching and problem solving |
| 9 | Coordinating actions across groups |
| 10 | Assessing the results of improvement efforts |
| 11 | Developing new plans to solve continuing problems or to achieve new results |

The first is to develop and clarify goals. Developing clear goals is essential to specify measurable outcomes or results that are the target of proposed changes. If you do not know where you are going, it is hard to develop plans for the trip or to know whether you have arrived at the destination.

A second technical activity is the collection and analysis of existing school outcomes. To what extent is the school achieving its objectives? In many improvement efforts the focus is on student achievement data. Schools collect norm-referenced, criterion-referenced, or authentic performance data. These are examined to determine whether all students—irrespective of level, race, or background—are achieving at desired levels. These performance data provide a picture of where the school has succeeded or fallen short. These data are used to set performance targets.

Next, the principal develops a planning team. In some cases this involves a formal and structured group team while in other cases the team is composed of volunteers who meet informally. Without some form of planning team, group decision-making and the identification of potential solutions for closing the gaps between actual and desired performance is a difficult, if not impossible, undertaking.

Working jointly with the planning team, the principal develops activities that will achieve improvement goals. The planning process identifies the purpose and mission of the school, assesses school outcome data, and develops a set of programs or activities designed to solve the problems. School improvement teams often meet over the summer or before school to begin to develop a plan. Plans typically specify goals, activities, time lines, resources, and personnel and outline ways to evaluate progress.

All plans need to be put into action. Often improvement plans have not succeeded because they fail to address the challenges and obstacles that arise as plans move from drawing board to working practices. Force field analysis (Lewin 1952) can often help to identify the shortfall between reach and grasp. Strategies for implementing the plan are just as important as the plan itself.

Successful improvement also requires resources, including time and materials for staff development and training. The principal and others gather resources and see that they are provided at the right time. Successful principals are entrepreneurs. They seek resources wherever they can find them—from central office, local businesses, or within their own budgets.

All changes require new skills. Technically oriented principals provide needed training and staff development. They identify the skills and capacities that are needed and find ways to develop these, including the provision of time and resources that

BEST COPY AVAILABLE

are needed. Without adequate training, all changes falter and die on the vine. Principals provide coaching and assist in problem-solving. They provide direct assistance or find others to help so that new skills can be mastered and reviewed on an ongoing basis.

Principals also coordinate actions across groups (Louis and Miles 1990). Schools are interlocking systems of activity. When one individual or group changes, it usually impacts others. Unless the changes are coordinated among various participants and groups, confusion and conflict are inevitable. Without coordination, school improvement often grinds to a halt.

Assessing the results of improvement efforts is an important part of technical change. Evaluation identifies whether or not the change effort has achieved its desired ends. Good evaluation taps multiple data sources. Comprehensive student assessments include achievement data, surveys of staff's, students', and parents' perceptions, or other measures of impact. Evaluation identifies how well the improvement effort has reshaped school problems or practices. The technical improvement cycle begins anew with the development of plans to achieve emerging goals that previous solutions did not accomplish. Or, other goals or purposes are intended in the process of comparing reality to desirable targets.

Each of these technical activities is an important link in the interlocking chain of school improvement. Unlike the connections of a chain, however, actions are not linear and reverberate back and forth in the interlocking motion of a human system.

## Symbolic Leadership

But the interlocking technical motions of a human system often occur on the surface. Underneath the rational membrane of a school, there is a swirling vortex of organic, existential activity. This deep flow has very little to do with anything rational or technical and a lot to do with traditional and spiritual values and beliefs. Successful change requires attention to the existential underbelly of the educational enterprise (see Figure 2).

Schools are not blank slates awaiting the artistry of a new principal. Most schools have been working to improve themselves for decades. Each month or year brings a new reform agenda—often before the preceding efforts were given a chance or even tried. A school's history is a very powerful force in shaping the present or future.

Before principals or others launch new programs, they first need to understand why things are the way they are and how they got that way. Principals listen to stories teachers tell about what in the past they consider important and hear descriptions of previous improvement efforts in the school. Without a deeper understanding of the history and existing beliefs of the school, principals are placed at a disadvantage as they try to move ahead.

Figure 2

SYMBOLIC ASPECTS OF IMPROVEMENT

| | |
|---|---|
| 1 | *Reading and understanding the school's history, norms, values, and traditions* |
| 2 | *Assessing the connection between existing cultural patterns and a vision of what the school might become* |
| 3 | *Creating transition events to help people let go of old norms and begin to grasp the meaning of new values and practices* |
| 4 | *Working with the informal network to develop support for emerging cultural patterns* |
| 5 | *Transforming the culture through leadership and symbolic activity* |
| 6 | *Encouraging periodic events to assess, enhance, and deepen cultural development* |
| 7 | *Building a commitment to continuous cultural evolution* |

In addition to looking backward at historical events, principals can develop a sense of where the school needs to go. Contrary to the popular notion that principals as prophets must deliver a vision, they are more like mediums who need to lend an ear to the dreams and wishes of teachers and others. Anyone who does a job has ideas about how it can be done better. A shared sense of a desirable future is the best way of shedding light on what needs to be done to improve educational experiences for young people. The principal's job is to weave individual visions into a shared dream and to help people see the gap between the reality of where the school is and the vision of where it needs to be.

Even with a shared vision as a guide, transforming a school is never as easy as it may first appear. Given that people agree on what needs to be done and share a vision, there is no guarantee that the dream will materialize. Cultural traditions and practices are not easy to change because people are more deeply attached to existing norms and values than they consciously realize. When the attachment is threatened or broken, they experience an unconscious sense of loss and either return to the past or embrace a meaningless future by losing themselves in frantic routine activity. A school soon falls into the old ways, spinning its wheels until it becomes stuck

BEST COPY AVAILABLE

somewhere between the old and new. Griping in the lounge about change is often a symptom that people are grieving and need to mourn.

Successful improvement requires that people let go of old ways before they can grab onto new ones. It is helpful to have opportunities to air conflicts and disagreements, to mourn and figuratively bury old practices and commemorate them in a visible museum—a display of the way we once were. Transition events are never easy and the pain can be lessened by reviving and revitalizing patterns and practices from the past. These are often laid aside even though they have future merit (giving old textbooks to a needy school is one example). But every change is a loss and all losses need to be recognized and dealt with through shared experiences that uncover and repair the deep issues that inevitably arise when cultural patterns and practices are reviewed, revised, or revisited. This is especially true of students and parents who are often excluded form the grieving and grabbing-on process. Their school changes and programs are lost without a grieving ceremony.

In the process of improvement, principals stay in close touch with the network of cultural players. People in these informal roles have considerable information and power. Normally they are seen as adversaries when they can be cultivated as allies or used as important foils. Priests or priestesses are the protectors of traditions. They preside over important cultural events, and are confidants and confessors to everyone. (In one school the reading specialist is talked to by everyone whenever a concern arises.) Although these spiritual leaders are often the most powerful blockers, they can become allies—if they see that the changes are reviving old values, if they believe that the new directions are necessary for survival, or that the changes will service the flock's spiritual needs better than the status quo.

Through their tales, school storytellers spread and dramatize a school's lore. Stories help to infuse the enterprise with meaning and purpose and to keep historical values and beliefs current and full of fresh passion. Storytellers can become allies if they are courted and provided with raw material that can be woven and crafted into a new story. Improving a school often means changing or revising existing stories. Such is the gift of storytellers. They are often the only ones that can reshape the lore. Seasoned teachers have many stories of past changes—both successful and unsuccessful.

Gossips keep everyone current. When change is in the air, the informal switchboard flows with new rumors and tidbits. Often, gossip is anti-change. Even then, it is important for a principal to know what's in the grapevine. Even better, the principal can rely on the gossip of change and improvement to spread the word. The further removed from the source, the more distorted gossip becomes. Working with, rather than isolating or fighting, the network of cultural players is an essential part of successful school improvement.

As cultural transformation proceeds, a principal can shape its direction in several ways. She can anoint and celebrate new heroes or heroes who exemplify and embody new values. By providing new phrases, stories, and metaphors the principal can provide linguistic support for emerging new patterns and practices. By encouraging new rituals and ceremonies, the principal can ensure that the community bonds to new values as well as to each other.

When events occur, principals take on many roles. A principal herself can exemplify new practices and patterns. Through dress, demeanor, costumes, and artifacts, principals signal what they value and appreciate. Successful school improvement requires principals who realize that their words and deeds must echo and model new values and beliefs.

At various stages of cultural transformation, a school needs time to reflect. Where did we begin? Where are we now? Are we on the right track? Where do we need to stretch to get to where we all want to go? Often things seem worse because people's expectations have risen far above where they once were. Sometimes people seek an arena to gripe or an occasion to grieve a little more. More often, people enjoy a pat on the back, a shot in the arm, and a chance to celebrate any progress that has been made. In the Beaverton Public Schools, an annual "Pits and Berries Night" gives everyone a chance to tell stories about the year's low points and high moments. Through the stories, people look back and then join forces to rev-up for the coming year.

Educational cultures find ways to recognize performance. Too often, school improvement progress is evaluated numerically when expressive approaches yield a better shot of what's working, what isn't, and where we go from here.

Finally, it is helpful to have continuous improvement as one of a school's core values. As soon as you rest on your laurels, you start to fall behind. Successful organizations take time to look back at how far they have come but then quickly look ahead to see how far they need to go on an endless journey of getting better in response to new challenges.

## REFERENCES

Lewin, K. (1952). "Group Decision and Social Change." In *Readings in Social Psychology*, edited by G.E. Swanson and E.L. Hartley. New York: Henry Holt and Company.

Louis, K.S., and M.B. Miles. (1990). *Improving the Urban High School: What Works and Why*. New York: Teachers College, Columbia University.

Wilson, B. and W. Firestone. (1987). "The Principal and Instruction: Combining Bureaucratic and Culture Linkages." *Educational Leadership* 45, 1: 18-29.

# VIII-6

# What We've Learned: Reflections on Implementing Performance Assessment

TIM WESTERBERG
LITTLETON HIGH SCHOOL, LITTLETON, COLORADO

Littleton High School, a comprehensive public high school in suburban Denver, Colorado, is a pioneer in the field of performance-based education. Since the fall of 1991, members of the class of 1995 and each succeeding class have been proving that they know and can do the things the school values as outlined initially in 19 performance-based graduation requirements and more recently in 28 course-based performance assessments called essential tasks. Students work with academic advisors, and specific performance standards and expectations are communicated to students and parents by teachers and through published course descriptions. Decisions about curriculum, staff development, budget, and other school "structures" are based on our outcomes. Littleton High School's asessment work is described in the April 1990 and May 1991 issues of *Educational Leadership* and has achieved a fair amount of notoriety.

As Littleton High School completes the seventh year of research, development, and implementation of its nationally recognized school restructuring program, certain realities of the change process have become evident. The 15 reflections on change at Littleton High School presented below are based, in part, on the work of Fullan and Miles (1992) and are offered here to others who may benefit from this information as they move through school restructuring and toward the implementation of performance assessment in their own schools.

## Reflections on Change

### 1. Change Is Resource Hungry

The change process in which we have been involved during the last seven years has been tremendously demanding in terms of both human and financial resources. Redesigning a system of high school education that has remained virtually unchanged for one hundred years is very time-consuming. Schools cannot restructure education completely on the backs of volunteers. Administrators must periodically release teachers and other school personnel from their teaching duties and provide them with summer stipends to redesign school beliefs, goals, and structures. The development of performance assessments alone consumed over 1,500 hours of paid staff time at Littleton High School during each of two summers. The toll on the energy levels of the school's human resources is considerable. Burnout of key personnel is a constant threat, even when commitment is high. Private and public financial assistance are necessary to fuel the engines of change. The Hewlett-Packard Company's practice of awarding incentive grants to schools, both from its corporate headquarters in Palo Alto, California, and from many of its individual plant sites throughout the country, serves as a model for the private sector. School districts, intermediate units, and states must set aside funds for innovative projects, and even the federal government has a legitimate goal in funding research and development in school restructuring.

### 2. Everything We Have Done Has Taken Longer Than We Thought It Would

For planning purposes, an educational change agent would do well to double the amount of time estimated for completing specific components of the restructuring project. Our experiences at Littleton High School have repeatedly suggested that we are not realistic when estimating the amount of time it will take to complete specific tasks. For example, what we had estimated as a one-year project—the development of performance assessments designed to evaluate student achievement on our performance-based graduation requirements—took two full years to complete. A similar "time warp" occurred in the development of the graduation requirements themselves. The results of our planning naiveté have been the predictable frustration and disillusionment that come with not meeting established deadlines.

### 3. Change Is Loaded with Uncertainty

One of the reasons it takes longer than anticipated to accomplish many school restructuring tasks is that the school transformation process is filled with uncertainty. Unanticipated problems, questions, hurdles, and delays continually crop up. We have found that school restructuring is not a linear process. Planning must be fluid and flexible. Uncertainty creates morale problems for those members of the school staff and community who think in a linear fashion and who like to have all questions answered up front. People of this disposition must receive constant reassurance. Restructuring works best in a community of learners, a community in which key players have a disposition toward learning new things and solving problems throughout the change process.

### 4. Restructuring Is Systemic Change

The school restructuring process at Littleton High School began with the end of the educational process clearly in sight. That is, we started our process by defining, in performance-based terms, exactly what it is that we want students to know and be able to do when they graduate from our school. Beginning with the goals of the process clearly in mind has resulted in significant changes in the structures of our school. Curricular changes, staff development needs, changes in the master schedule, and decisions about the allocation of resources are based on our educational goals. Restructuring is not tinkering with some aspect of the school environment. Instead, it is a process of redesigning the social order of the school to enable the organization to meet its goals. Peter Senge's *The Fifth Discipline* (1990) is an excellent resource on the systems thinking needed for successful restructuring based on performance assessment.

### 5. Implementation of Performance Assessment at the Local Level Is Heavily Influenced by the Opinions of the NPU (National Psychometricians Union)

Actually, no such union exists but you couldn't convince the Littleton High School faculty of that. Decisions involving performance assessment made on the basis of "what's best for kids" have had to be altered, or changed in some cases, in favor of what a relatively small number of nationally recognized "experts" tell us is technically and legally defensible. To confuse matters further, we are required to modify locally developed, high-stakes performance assessments to meet the technical standards recommended by psychometricians, but we are not required to modify traditional teacher-made classroom assessments to meet these standards (a task that would take most schools years to complete; see Reflection 8 for more on this topic). Those considering implementing a system of performance assessment in their school, especially if they intend to use the system to make high-stakes decisions such as eligibility for graduation, should be cognizant of all technical and legal ramifications. Have you hugged your psychometrician today?

### 6. Problems Are Our Friends; Expect a Lot of Friends

Significant change produces critics, and we can either shut critics out of the change process or use them to strengthen it. The restructuring effort at Littleton High School has its share of critics, some well-intentioned and some not. There is a temptation to slough off criticism as the reaction of extremists and not representative of the true status of the project. We have found that the best approach is to bring the critics into the circle of influence and to get their questions and objections "on the table" for full examination. Critics often have legitimate arguments, and healthy skepticism can help unearth potential problems before they threaten the change effort.

### 7. Everybody Wants Change Until You Do It

Given the public cries for change in education across this nation since 1983, we assumed that our school community would enthusiastically embrace our plans to turn the present educational system upside down. Wrong! We should have taken a clue from the annual Gallup Poll on education. Each year the respondents in that poll give low marks to education in general but say that the schools their children attend are doing just fine. Unless you are restructuring a school already in crisis, be prepared for the "if it ain't broke, don't fix it" mentality. Parents will question "experimenting" with children, and students will admit that raising standards is a good idea, but for the next class. The line from the old blues song, "Everybody wants to go to heaven but nobody wants to die," applies to school transformation at the local level.

### 8. The New System Is Always Held to Higher Standards Than the Old System Was

Responding to the skeptic's concern about a perceived inadequacy in the restructured system with the statement, "The old system has that problem too," is seldom satisfactory. For example, people sometimes ask how we plan to ensure that all teachers score student performances the same way (interrater reliability) when, in fact, no such assurance was built into our old system. Most people would rather live with a familiar pain than trade it for the promise of an unknown pleasure. In education, and in other high-stakes matters, a new system always is, and probably always should be, held to higher standards than the familiar system was.

### 9. Effective Communication Is Our Biggest Problem

In spite of a fairly substantial and well-supported public relations campaign, we are continually burdened by both internal and external communications problems. After dozens of memoranda and faculty meetings, staff members still profess to be totally unaware of important directions and decisions. At times rumor runs rampant in the community. Regular newsletters are important, but don't expect everyone to read them. Neighborhood coffees, open forums, and a lot of one-on-one interaction are necessary to offset this reality.

### 10. Broad-Based Community Support and Involvement Are Essential

We can ameliorate each of the above realities to some extent by having representatives from a cross-section of the community involved in the school change process. Our parents, students, and business partners have been invaluable in communicating the need for change and the particular changes underway at Littleton High School to our community. Each of these groups has helped to secure resources for the project as well. Uncertainty and fear are greatly diminished among key groups in the community when members of those groups speak with authority in favor of the restructuring effort.

### 11. For All the Talk About School-Centered Decision Making, Districtwide Politics Still Determine What Gets Done

In the fall of 1993, the Littleton Public Schools experienced one of the most contentious elections in the district's history. Three back-to-basics candidates, running as a slate on a platform that promised to stop many of the reforms underway in the district, won seats on a five-member board of education. In spite of what we believe was a fairly effective communications plan in our local attendance area, controversy brewing over issues elsewhere in the district (high school restructuring, middle level education, whole language instruction, and shared decision making, for example) produced enough voter dissatisfaction to force a change of direction.

The new board has not totally dismantled our restructuring efforts, but significant changes have occurred. What we learned, in part, is that depending on a single-issue, local school public relations effort to maintain a base of political support can be a mistake.

### 12. Central Office and Board of Education Support Are Essential

The school transformation process underway at Littleton High School would not have advanced to its present state without the nurturing and protective environment that the superintendents and the Boards of Education established in the 1980s and early 1990s. During that time, Littleton Public Schools encouraged educators to be creative, take risks, and stretch traditional boundaries. Just as important, a visionary strategic plan and other board policies legitimized and protected our restructuring efforts at Littleton High School. Our survival in the future will also depend, to a large degree, on support from the superintendent and the Board of Education. Bottom-up innovation and risk-taking depend upon top-down protection and support.

### 13. Conceptual Clarity Is Important to the Local Implementation of Change

Alternative assessment, authentic assessment, comprehensive assessment, direct assessment, practical testing.... Which is it? Worthen (1993) makes the point that advancement in any field depends upon the possession of "universal" language. Some of the communications difficulties we have experienced at Littleton High School, and a possible factor contributing to the voter dissatisfaction discussed above, is the use of language that is confusing to the publics we serve. Developing a glossary of terms early on in the change process, one that is relatively free of "educationese," is well-advised.

### 14. Powerful Ideas Withstand the Winds of Change

As 1995 drew to a close, we were not yet sure what further changes might be made in our performance-based assessment programs as a result of continuing political upheaval in the district. What was obvious, however, was that the skills, knowledge, ideas, conviction, and commitments gained and made by the Littleton High School staff would carry on into the future. There is a great comfort in the fact that powerful ideas can withstand changing times.

### 15. Restructuring Means High Excitement and High Anxiety

Carl Glickman at the University of Georgia first coined this phrase describing school restructuring. Pioneering new developments in the high-stakes arena of public education produces anxiety. There are times at Littleton High School when we wonder if we have taken on more than we can handle, if we are in over our heads, if we are doing the right thing. But true school restructuring can also be tremendously exciting. Littleton High School, under the influence of performance-driven restructuring, is an intellectually exciting place to be. Members of the school community are discussing, debating, and deciding important issues. And isn't that the way schools ought to be?

## REFERENCES AND RESOURCES

Brickley, D., and T. Westerberg. (April 1990). "Restructuring a Comprehensive High School." *Educational Leadership* 47, 7: 28-31.

Fullan, M., and M. Miles. (June 1992). "Getting Reform Right: What Works and What Doesn't." *Phi Delta Kappan* 73, 10: 745-752.

Senge, P.M. (1990). *The Fifth Discipline: The Art and Practice of the Learning Organization*. New York: Doubleday/Currency.

Westerberg, T.R., and D. Brickley. (May 1991). "Restructuring a Comprehensive High School." *Educational Leadership* 48, 8: 23-26.

Worthen, B.R. (February 1993). "Critical Issues That Will Determine the Future of Alternative Assessment." *Phi Delta Kappan* 74, 6: 444-457.

# VIII-7

# Overcoming Barriers to Educational Change

Michael G. Fullan
University of Toronto, Canada

Practical research over the past 20 years has provided us with a growing sense of the do's and don'ts that make for successful change (Fullan 1991). Following a few words about the nature of educational change, I will address the questions: What are the key barriers to achieving educational change? And what insights and strategies are likely to overcome the barriers?

## What Is Educational Change?

The foremost thing we have learned, which seems obvious but is usually misunderstood, is that change stands or falls on the motivations and skills of teachers. When one attempts to define the essence of change, the focus inevitably zeroes in on the teacher. We can ask what types of things have changed when a given instructional innovation or reform successfully takes place. Specifically, the core of educational change concerns alterations in teaching in at least three components: material/structures, practices/behaviors/skills, beliefs/understanding.

As we turn to barriers and strategies, the overarching question is what kinds of factors or conditions inhibit or help teachers and others to work through a successful process of change.

## Barriers to Change

There are many different ways to discuss barriers to educational change. I will focus on six problems closely linked to change because they are amenable to action and, if addressed effectively, can improve our success rate. But first we must understand them. Six basic change-related problems are: Overload; Complexity; (In)Compatibility; (Lack of) Capability; (Limited) Resources; Poor Change Strategies. These six problems interact in a downward spiral.

### Overload

It would be hard enough if all schools had to do was to implement one innovation at a time. But schools are in the business of attempting to cope with *multiple innovations simultaneously.* Overload is caused by a combination of too many pressing needs and a political system that generates far more innovations than can be managed. In fact, it is probably closer to the truth to observe that the main problem in North American education today is not the absence of or resistance to innovation but rather the

---

First published as "Overcoming Barriers to Educational Changing" in Changing Schools: Insights (Washington, D.C.: Office of Policy and Planning, U.S. Department of Education, 1992), pp. 11-20.

presence of too many ad hoc, fragmented, uncoordinated changes. Any strategy for reform that is going to work must address the problem of overload and fragmentation of effort.

### Complexity

A second problem is that initiators of change have underestimated how complex serious change really is. It is complex in two senses. The first, and most obvious, relates to the development of new skills and the reality of personal anxieties and uncertainties that characterize the process of change for individuals—all the more complicated when these problems have to be coordinated in social settings to achieve *shared* solutions. Second, and less obvious, is that, in more radical reforms, we literally do not know what the solution will or should be at the early stages. In other words, the solution has not yet been developed; the exact nature of the reform is not known.

### (In)Compatibility

Another major problem concerns the philosophical compatibility of a reform with the personal philosophies of the educators who are to implement it. A teacher who believes in Direct Instruction will have great difficulty accepting Whole Language as the teaching approach to use. Elementary school teachers endorse mixed ability classes more often than secondary school teachers do, because of differing philosophical beliefs about the whole child. So, the question of basic compatibility with a reform must be confronted. But a point should be raised here. When people resist or do not implement an innovation, they may be doing so out of uncertainty rather than basic incompatibility. One can see how this problem also interacts with complexity and capability. If a problem is complex (unknown), and if we do not have the capacity (skills) to address it, it may surface as a matter of incompatibility: we are against the change.

Thus, (in)compatibility is a problem, but will be treated here as a variable. As people become clearer and more skilled in something and see that it has positive effects, the compatibility problem may vanish.

### (Lack of) Capability

Because successful change involves learning how to do something new, it follows that skill and know-how are crucial. Probably the major barrier in dealing with complex changes is the development of the technical capacity needed to master the new situation or program. At the early stages of complex change, the costs (time, energy, and psychological costs) are high and the benefits (progress, impact) are low. Thus, when people are having difficulty or avoiding needed change, it may have just as much to do with capacity as it does with compatibility or beliefs. It is necessary, then, for any reform effort to include capacity building (of the individual and the organization) as a foundation strategy. Developing skills and know-how must be seen as a natural and intrinsic part of any serious change effort.

### (Limited) Resources

All of the processes discussed thus far require some additional resources—time, money, ideas, expertise—if the change strategy is to be successful. In Miles's terms, successful change projects are resource hungry (Louis and Miles 1990, Fullan and Miles 1992), which does not and cannot always mean money. There are many ways that effective schools create additional time and access to resources without adding to the budget. However, some significant additional resources, especially in the form of time and expertise, is required for substantial reform.

A second and more subtle aspect of this barrier is what one might call "the evaporation of resources." Large-scale, politically backed reforms nearly always have millions of dollars attached to them when they are announced by political leaders. It is not a reflection on the integrity of change sponsors to say that teachers and principals at the ground level often do not see significant new resources. This is all the more galling when those who are struggling with reform hear public statements referring to the vast amounts of money allocated to that reform.

As others have observed, high-profile reference to reform initiatives and associated resources is as much a political and symbolic act as it is a substantial implementation act. Successful change projects convert this political energy into resources at the level of school reform where principals and teachers both experience and perceive that additional investments needed for change have been made.

### Poor Change Strategies

This is something of a metabarrier. The question is, do the strategies of change normally used address the five major barriers just discussed? There is overwhelming evidence that they do not (Fullan 1991, Fullan and Miles 1992). New legislation, accountability and monitoring schemes, one-shot professional development (no matter how inspiring), allocation of money, creation of new roles, wishful thinking, exhortation to leaders and teachers, and appeals about the ills of society and the terrible needs of at-risk children do not in themselves form effective strategies for reform. At best, they inspire people to start or want to start. At worst, they create cynicism as people see reform initiatives come and go, and begin to believe that reform is political and marginal to their everyday demands and interests.

The overall conclusion, then, is that if a reform initiative is to have any chance for success, it must incorporate systematic and continuous strategies that confront and address (not necessarily resolve) the five major barriers outlined above. Up to this point, the change strategies themselves have created an additional barrier because, not only have they not worked, but they have made matters worse.

## Strategies for Change

In this section I will refer to some more elaborate sources on strategies for change, identify the main components that a set of strategies should contain in order to address the six obstacles discussed earlier, and discuss the time frame required for success. These recommendations are based on recent research on the change process.

First, there is no single answer to the question of what strategies are best, because it depends on the particular context and role or vantage point. To a certain extent, strategy advice must be *particularized*. We have done this elsewhere with respect to role—of teachers, parents, principals, students, district administrators, and governments (Fullan 1991); to themes such as establishing collaborative work cultures (Fullan and Hargreaves 1991); and to guiding propositions for success (Fullan and Miles 1992).

Second, given our experience and our knowledge of the change process, what strategies do we recommend? The six essential components are to develop a describable change strategy; to engage in evolutionary planning and consensus building; to employ technical assistance for all; to focus on building collaborative work cultures in the school and community; to work on school-district and school-external relationships; and to establish an inquiring, problem-solving, monitoring mind-set.

### 1. To Develop a Describable Change Strategy

Because previous strategies have been ineffective, crude, and unstated, it is necessary to make the change strategy *explicit* for all. This does not mean rigid strategy, but it does mean that the main components of the strategy are formulated, debated, and agreed upon. A relevant example is the Institute for Responsive Educators' (IRE) League of Schools Reaching Out—a national strategy involving 65 schools "to show how family/community/school partnerships can contribute significantly to school restructuring aimed directly at increasing the academic and social success of all children, especially those society labels 'at risk'" (Davies 1991). IRE explicitly states its strategy for reform, involving a national network of schools using seven core strategy components: ideology (a conceptual and value-based focus); felt need (participants who feel a need for substantial change); an action plan (in this case, parent centers, minigrants, etc.); on-site help (skilled and committed field facilitators); recognition (recognition and visibility to participants); discretionary money; and "loving critics" (respect and useful criticism from third-party reform agents).

The point is not to accept this exact list, but it does represent the *type* of strategies needed. The change process is subtle, however. Although in the Learning Consortium we started with a guiding ideology and some felt need (points 1 and 2), the process of change was one of mobilizing greater and greater consensus and felt need. In other words, a program doesn't start with crystal-clear ideology and widespread consensual need but works toward it.

We did have multifaceted action plans that involved many components. These included summer institutes (5 residential days of training with school teams of principals and teachers and central office support staff); onsite follow-up; nurturing the informal/collaborative culture in schools; leadership training; provision of staff development and other resources; a spirit of inquiry that encompasses minigrants, collaborative research between practitioners and academics, external case studies (loving critics), and formal evaluation; and use of the bureaucracy to buttress the change efforts—teacher appraisal, site-based management policies, leadership training schemes, redesigning preservice teacher education, and new induction programs. Onsite help, recognition, and discretionary money were incorporated into these plans.

Thus, the first requirement is to work on an overriding strategy with named components. The next five components contain additional ideas for what such a strategy should incorporate.

### 2. To Engage in Evolutionary Planning and Consensus Building

The best discussion of these twin processes is in Louis and Miles (1990, Chapters 8 and 9). The idea is to establish some guiding principles but to begin to work on them in practice as the route to sorting out and developing vision and consensus. In our experience it is a mistake to fix on a vision too early. It is better to identify promising themes (e.g., parent centers in schools, partnerships with other social agencies in the community and with businesses, student engagement in learning, building collaborative work cultures), to work on them, and then to forge consensus based on experience and growing expertise. The remaining four components directly feed into and support evolutionary planning and consensus building.

### 3. To Employ Technical Assistance for All

A caveat: Item 4 establishes a great deal of day-to-day assistance in collaborative schools, so we should not think of technical assistance as confined to designated workshops or inservice training. Nonetheless, ongoing assistance involving outside facilitators is necessary. For a staff of 30, an estimate is some 30 person-days allocated to small group or team inservice training (e.g., two teams of five receiving a total of three days per staff). Most of these days involve outside facilitators (school district or external facilitators). I do not believe that any useful formula can be established. The lesson, however, is clear: designated, ongoing technical assistance of high quality is needed year after year to bring about serious change. Similar inservice is required (often as part of teams) for administrators, parents or other community participants, and students if they are part of a change project.

Finally, technical assistance is more than direct inservice training. It also means materials, ideas, time—that is, resources more broadly defined. Louis and Miles (1990) suggest that a "floor" of $50,000 to $100,000 a year for several years is needed for urban high school improvement, but I would hesitate to attach a particular dollar figure in the abstract.

### 4. To Focus on Building Collaborative Work Cultures

We have worked extensively on building collaborative work cultures focusing on norms and practices of *collegiality* (peer coaching, mentoring, giving and receiving assistance); *continuous improvement* (continually seeking and testing new ideas that work); and *restructuring* (developing a timetable and resources) in support of collaboration among teachers and principals (Fullan and Hargreaves 1991). This is very much a generative resource in two senses. First, these practices create, both informally and formally, what is tantamount to a great deal of day-to-day inservice training and assistance. Learning among staff permeates the organization—see also Rosenholtz's (1989) "stuck" and "moving" schools. Second, collaborative work cultures provide the glue for technical assistance to be sought, to be used effectively, and ultimately to have an impact. No amount of technical assistance by itself (even if it is 20 days a year for the entire school staff) will have an impact unless it is linked to collaborative working relationships. When it does occur in these situations, its impact multiplies.

### 5. To Work on School-District and School-External Relationships

Although the school is the center of change, it is a mistake to focus only on school development. Individual schools can develop effectively for short periods of time despite the district, but they cannot continue to be effective despite the district. Eventually, hiring decisions, transfers, allocation of resources, policies, etc., affect the school, for better or worse.

Therefore, any school development strategy must attend to school district development. This is crucial to any long-term success. In our experience it is most effective when staff at the district level, backed by research ideas, actively work on reorganizing district practices so that they both support and press for school development. This means mission and vision development, hiring and promotion procedures, teacher and administrator appraisal schemes, site-based planning requirements, staff development allocation and policy, assessment and evaluation, and organizational structure at the district level. These normally segmented and fragmented components of a district's operation can be coordinated.

The general message is: In addition to the networking among schools involved in a common reform strategy, there must be a component that works on the relationship of the schools to their districts and to other schools in their district not involved in the project. Without this component, project schools will not continue to develop in the long run, even if they do in the short run. Moreover, the strategy recommended here actually uses school development projects as a catalyst or means of stimulating district (and thus other school) development.

### 6. To Establish an Inquiring, Problem-solving, Monitoring Mind-Set

One major implication arising from the set of six barriers to change discussed earlier is that the pursuit of complex change requires a risk-taking, problem-solving, inquiry-feedback mind-set and corresponding mechanisms for action. We have established a multifaceted approach. First, we have committed to and engage in constant inquiry, using as many opportunities as we can find to conduct our own and third-party studies and to disseminate the findings. These activities stimulate participants to think and question, expose them to other ideas in the research literature, and provide recognition through dissemination.

Second, we have established mechanisms or regular forums for monitoring and problem solving, such as a steering group that meets, communicates, and solves problems regularly. Each school district, and each school within districts, has mechanisms for monitoring, such as school growth plans and assessment of progress. In some schools the practices are increasingly well-established, in others not so. As Louis and Miles (1990) found, schools differ in their problem-solving capacity. Some ignore or don't identify problems, others engage in superficial coping, still others in deep coping and problem solving. The implication is that training in monitoring and problem solving needs to be part of school development strategies, *provided that* it is integrated with the other strategy components.

Third, we have done less systematic formal assessment, so we have less advice to offer here. Ontario does not have a well-developed assessment system for students, and there are simply no standard data available. We have a good deal of qualitative data, including some systematically gathered, on implementation of new ideas, overall system measures of teacher collaboration, principal effectiveness using indicators from the research literature, and the like. After three years, we are only now turning our attention to student assessment and are currently considering a project on this aspect as we enter our second three-year phase. We agree that more quantitative indicators of teacher and student progress should be incorporated earlier, although we caution that if done prematurely, prescriptively, or independent of a spirit of inquiry it will actually inhibit rather than stimulate further development. As much as possible people within the project must be committed to seeking measures of impact.

## Conclusion

In reflecting on the change process, one should ask to what extent does the approach selected actually address, as systematically as can be done, the six barriers to change identified earlier. Successful change projects are efforts in *mobilization* and *positive contagion*. Ownership grows throughout successful reform projects, not something that is established in advance.

Finally, let me talk about time frame. For single innovations such as implementing a districtwide reading or science program, it will take approximately three years, assuming that good strategies are being used. For more substantial restructuring reforms our estimate is five to 10 years, depending on the scope of the reform (and again assuming that effective strategies are used). This does not mean the absence of early progress. Indeed, recognition and some early success are critical. But thorough change, well-institutionalized, must be conceived in terms of sets of years. We made a three-year initial commitment, with a provision for review and continuation for a second three-year phase. In any case, more than four or five years in total will likely be needed.

There is much more to the subtleties of the change process. What has become clear is that both good ideas and good strategies must be part of the same mix. The latter will only happen if we conduct and learn from serious change initiatives that attempt to incorporate both reform ideas and reform strategies.

## REFERENCES

Davies, D. (1991). "Testing a Strategy for Reform: The League of Schools Reaching Out." Paper presented at the American Educational Research Association annual meeting.

Fullan, M., with S. Steigelbauer. (1991). *The New Meaning of Educational Change.* Toronto: OISE Press; New York: Teachers College Press.

Fullan, M., and M. Miles. (June 1992). "Getting Reform Right: What Works and What Doesn't." *Phi Delta Kappan* 73, 10: 744-752.

Fullan, M., and A. Hargreaves. (1991). *What's Worth Fighting For? Working Together for Your School.* Andover: The Regional Laboratory; Toronto: Ontario Public School Teachers Federation.

Louis, K., and M. Miles. (1990). *Improving the Urban High School.* New York: Teachers College Press.

Rosenholtz, S. (1989). *Teachers' Workplace: The Social Organization of Schools.* New York: Longman.

# VIII-8

# Leadership and the Process of Changing Practice

Shirley M. Hord
Southwest Educational Development Laboratory

New methods and models of performance assessment such as those described in earlier chapters represent a significant change in the way teachers have typically monitored and evaluated students' learning and development. Investing attention, time, and resources in the processes of developing, adopting, and implementing new methods is also a significant departure from the manner in which schools have managed such changes in the past. In this paper we explore several issues vital to the successful installation of a change such as performance assessment in a school or school system. These issues include: What approaches do educators typically use to bring about change, and why don't they work? Why are leaders needed for successful change? Who are such leaders? What strategies do these leaders use to increase the success of change efforts? These questions are discussed as they relate to translating the vision of new authentic or performance assessment practices into their actual use with students in classrooms.

## Assumptions About Change

In the late sixties and early seventies, Chin and Benne (1969) and Havelock (1971) articulated several models of change, representing theories of how change works in organizations. The approaches of these models remain popular today, and two from Chin and Benne warrant a brief review.

### Empirical-Rational

The assumptions of this approach are that people are rational and always willing to take on new ideas, such as performance assessment, for improving student growth and learning. A further assumption is that, if a school system gives this well-intentioned receiving group a well-tested idea, program, or practice, the new program will find its way into classroom practice. In other words, if a *good idea* is given to *reasonable people*, they will put it into practice without further ado.

---

This paper was in large part abstracted from a monograph developed at the Southwest Educational Development Laboratory by Shirley M. Hord entitled "Facilitative Leadership: The Imperative for Change," 1992.

### Power-Coercive

A second approach to reform and school improvement is the use of power to leverage classroom change. In this case, an authority or strong power figure delivers a well-thought-out policy such as performance assessment to the receiving group. In the 1970s, such mandates typically came from the district administration; however in the 1980s and 1990s, as the goal and struggle for improving student outcomes has shifted to the state level, the state department of education has become an important player. In the power-coercive view of change, a *good idea* plus *pressure* on the receivers will achieve utilization.

While these two descriptions are admittedly simplistic, they represent what is commonly used to promote change in curriculum, instruction, and assessment. If we pay attention to change efforts that school systems initiate with enthusiasm but later allow to dwindle into oblivion, it is easy to observe the results of these approaches. After all, teachers and principals do not always have the expertise required to implement new curriculum programs and authentic assessment processes and use them effectively.

Fullan (1991) suggests that we have overestimated what people are able to do with new ideas. Time is also a factor: Don't teachers and principals have full-time jobs already? Even when a highly charismatic power figure delivers the finest new curricular approaches or instructional strategies to well-meaning and concerned educators, these approaches have not proven to be a recipe for successful change. This analysis is not meant to be discouraging, for schools and districts *do* change their ways of "doing" school and students *do* consequently benefit. How, then, does this happen?

### Leadership: The Key to Change

In studies of successful and unsuccessful school change, a factor consistently contributing to success has been the human interface—an individual or individuals who provided help to the implementors (Hord 1992). This help took many forms, but as Sashkin and Egermeier (1993) concluded, the "more personal assistance and continuing support from a skilled and knowledgeable local agent, the more likely that the innovation will be used" (p. 7) and real change will occur.

If the process requires a skilled and knowledgeable local agent, what do we know about these individuals? In the work on school change and improvement that the Southwest Educational Development Laboratory is now conducting, the local agent who is labeled the "facilitative leader" has become a major focus. Why use the word "leader"?

The extensive literature on educational leadership generally distinguishes between *management* of schools and *leadership* of schools. The literature views managers as rarely engaging in change; rather they arrange and organize existing practices and structures so that the organization functions smoothly and day-to-day routines are carried out—very important functions, to be sure. When they depart, managers leave things much as they found them (Hord 1992).

Leaders, however, focus on innovation and improvement. Leaders change and transform the organization according to a vision. They are changemakers, supporting and guiding the school or district to a new and more compelling vision to accomplish the goals of the district and school. Without leadership, the school's vision and goals are not realized. Without leadership, the use of new assessment practices is not likely to occur across a school. The outcome, then, of having leadership, whether represented by a single "hero" leader, a leadership team, or leadership widely shared across a school faculty and staff, is the attainment or achievement of desired change(s).

## Leaders—Anyone and Everyone

Who are the leaders who stimulate improvement? According to Murphy (1991), "There is no single key actor" (p. 32). Leadership, as noted above, is defined by function. It is not restricted to people occupying particular positions. Any person who can deliver the leadership function is a leader. Such persons can include principals, superintendents, and school board members. However, teachers, parents, and community members can also be significant educational leaders, as can central office consultants or specialists, external agency staff, and state department personnel. Students can act as leaders. Anyone who provides leadership, "the process of translating intentions into reality" (Block 1987, p. 98) can be a leader.

In a recent study of school board leadership, Hord and Reynolds-Gibbs (1993) found district leadership teams, composed of school board members and their superintendent, reaching directly into the community to solicit and suggest changes for the schools. In harmony with their district's mission, which focused on student outcomes, these boards interacted with their district constituents in developing a vision of improved effectiveness for students. The boards then acted to provide the resources and other support necessary to realize the vision—examples of providing leadership for change.

At the school level there is almost always more than one leader, though the literature most frequently points to the role of the principal. A leadership team that represents the various constituencies of the school and community is a structure currently recommended for school management and improvement purposes. The most recent literature focusing on workplace cultures that support and encourage continuous improvement recommends staffwide shared leadership that is grounded in a widely held vision of change and broadly shared decision making (Boyd 1992b). In such a culture, leadership is the responsibility of everyone, and the principal, no longer the "sage on stage," is rather the "guide on the side," facilitating and supporting the school's improvement efforts.

## Strategies for Successful Change

What, exactly, must the leaders do to ensure successful change efforts, such as the development and implementation of performance assessment? As Terrence Deal (1990) stresses, "Nothing will happen without leadership. From someone—or someplace—energy needs to be created, released, channeled, or mobilized to get the ball rolling in the right direction" (p. 4) and to keep it rolling. From a review of the literature on successful change, six leadership strategies or general categories of actions are identified and described below: create a context for change, develop and articulate the vision, plan and provide resources, invest in training and professional development, assess/monitor progress, and provide continuous assistance.

### Create a Context for Change

Educational professionals at all levels, students, and the community of parents, business, and lay citizens all influence school change efforts. Superintendents spend time and energy with business and the community-at-large managing issues external to the schools so that an internal context for change can flourish. They do this by tuning in to community concerns and by making frequent efforts to keep the community informed of upcoming changes. These superintendent actions are particularly important when related to a change in student assessment. They make sure that a harmonious environment pervades the district in order to nurture the internal creativity of the school staffs.

Internally, superintendents and their district level colleagues work actively to challenge administrators, teachers, and other staff to create innovative ideas and make suggestions for improvement. To develop this attitude toward change, they arrange meetings for staff to share their ideas, and they support staff's risk-taking activities by acknowledging that risk taking necessarily involves making mistakes. They declare consistently that mistakes lead to learning. Superintendents express the value of principals who take actions for change, not for the status quo (Paulu 1988). One set of actions principals and campus improvement teams take is to develop a school culture and context that supports change. Boyd (1992b, 1992a) reports several ways they do this, as well as the factors that characterize such a culture.

Boyd and Hord (1994) identify four functions of school context that contribute to successful change and that leaders nurture.

1. *Reduce isolation.* The principal and leadership team connect school staff by developing schedules and structures that bring staff together. They implement policies that foster collaboration, strive for effective communication, and develop collegial relationships among teachers, thus creating a sense of community in the school.

2. *Increase staff capacity.* The leadership of the school and district make policies that provide (a) greater autonomy and (b) staff development to school personnel. They ensure that resources are available and that staff have broad involvement in decision making.

3. *Provide a caring, productive environment.* A context conducive to change is one in which teacher attitudes toward schooling, students, and change are positive. Students are interested and engaged with learning. Community attitudes are supportive, and parents and community members serve as partners and allies. Leaders develop these conditions over a period of time, and they result in part from the first two functions.

4. *Promote increased quality.* When the leadership of a school instills a widely shared vision or sense of purpose, a norm of continuous critical inquiry, and a norm of continuous improvement, then action for school change becomes the standard operating procedure.

These four functions embody the "learning organization" espoused by Senge (1990). Unlike the empirical-rational and power-coercive approaches to change noted earlier in this paper, learning organizations thrive on a culture that values efforts at continuous improvement. Collegiality across the staff that nurtures reciprocal help and support, in addition to encouragement and pressure, is a key characteristic of learning organizations. Because structures and schedules provide an infrastructure for continuous conversation, individuals in learning organizations maintain consensus on their shared vision, and they participate broadly in decision making. Furthermore, leadership is inclusive and facilitation for changing practices is shared throughout the staff. This means that staff recognize the need for helping and supporting each other in implementing new assessment practices.

### Develop and Articulate the Vision

It is important to distinguish between *vision* and *mission*. We think of mission as the purpose for which the district or school exists and indicates the district/school's intended outcomes for students. Staff of North Central Regional Laboratory published an articulation of student outcomes that identified the following characteristics of successful learners (Jones and Fennimore 1990):

- *Knowledgeable.* Successful learners use their acquired knowledge to think fluently and with authority, having a strong sense of what they believe and why; they are problem solvers and evaluators of information, reflecting on and puzzling about information.
- *Self-determined.* Successful learners are motivated and feel effective in determining their own development; they believe in hard work and persevere with confidence when difficulties arise, assuming the world will provide opportunities and choices for their examination and action.
- *Strategic.* Successful students think about and control their own learning through use of strategies for learning various subjects; they plan, organize, monitor, and summarize their learning, orchestrating these learning activities.
- *Empathetic.* Successful students see themselves and the world through the eyes of others, examining their own beliefs and those of others and sharing experiences to increase understanding and appreciation; they recognize the value of communicating with others, as well as developing a range of interpersonal skills to interact with and develop appreciation of multiple cultures.

These articulated outcomes for students provide the basis for the mission. Vision, which is influenced and driven by the mission, refers to mental pictures of what the school or school district's programs and processes might look like in a changed and improved state—a preferred image of the future. An effective superintendent or principal does not act alone to create a vision and impose it; he or she builds a vision together with the participants of the school organization. In this way, "it becomes the common ground, the shared vision that compels all involved" (Méndez-Morse 1992).

The superintendent's initial vision is one factor that can drive the organization to achieve its next stage of evolution—through introducing new ideas and possibilities for improving the district's schools. The superintendent may introduce new student assessment procedures in this manner. Successful superintendents (those who provide effective programs for students and remain in their jobs) start with where the community and the staff are, work with these constituents collegially to study, review, and reflect on new ideas, and develop new visions and plans for improving the school system.

Regularly and frequently, the superintendent and all leaders at district and campus levels describe and promote the vision to the public through newsletters, radio, and TV interview shows; through weekly interactions at the Rotary and Lions' clubs; at the community's favorite coffee shop; and even at the football awards banquet. In addition, leaders spend a great deal of time personally articulating the vision and priorities with administrators, teachers, and all other staff. Boyd (1992b) advises, "A clear vision of the school when the change is successfully implemented and how implementation will occur needs to be developed among all in the school" (p. 51).

For any change, then, there should be a clear depiction of what it will look like when it is in operation: What roles will administrators, teachers, and students have relative to the new assessment practice or process? How will it differ from the current way of operating? Will parents have new roles or relationships in the new assessment scheme of things and if so, what will this look like? Change will not be realized if there is no clearly developed vision to guide its implementation.

Vision and goal setting establish the parameters for leaders' subsequent actions, giving them a clear image of their schools in order to set priorities (Manasse 1985). Leaders use the goals as a continuing source of motivation and planning (Blumberg and Greenfield 1980; Manasse 1984; MacPhail-Wilson and Guth 1983) and as a basis for providing clear, consistent, and well-communicated policy (McCurdy 1983).

### Plan and Provide Resources

The development of a plan that will achieve the new vision of assessment may evolve through interaction with participants or may be developed by the leader(s). Louis and Miles (1990) prefer the evolutionary approach. They suggest that leadership in change processes must involve planning, but effective planning should avoid a grandiose "blueprinting" approach in favor of a strong evolutionary character.

Both the change program and the school develop steadily, they say, when the goals for change and the shared vision of improvement for students drive the process. Data on the progress of the improvement effort suggest detours or new avenues; new capacities develop and permit more ambitious efforts than anyone had dreamed of previously. Evolutionary planning is not a hand-to-mouth approach but a coherent, intelligent adaptation based on direct experience with what is working in moving toward the new assessment vision and what is not.

The problem of resources—not only available dollars, but reallocations of time, personnel, existing materials and equipment, etc.—can improve implementation of the assessment plan. Although leaders are accustomed to thinking broadly about resources, they aren't always aware of what teachers need. Studies have shown that when leaders visit schools and classrooms and understand what the teachers are doing, resources available to the teachers gain relevance (Brandt 1987; Peterson 1984).

Fullan (1985) suggests that administrators place emphasis on such resources as released time for planning and training. Material resources are essential, but the time and energy demanded of people to plan, share, observe, and take action are at least of equal importance. In order to have the time and energy required for making positive changes (i.e., in assessment procedures), Anderson (1993) urges that implementers and their facilitating leaders must have other responsibilities decreased or eliminated in order to give attention to the "new." This is a knotty problem, probably best solved through staff brainstorming and collegial interaction.

Bridges (1991) addresses the issue in a practical, but slightly different, way. He maintains that there should be a formal identification of what has "ended" (old practices), in whose place the school or district will implement new practices. In introducing any change, it is useful to determine what to "discard" and what will "come on board."

According to Fullan (1991), "each school must be assisted by someone trained in supporting the endeavor.... [Such] assistance is directed toward facilitating and prodding the process" (p. 414). The *someone* is the "skilled and knowledgeable local agent." In Fullan's view, the school system should focus additional resources on developing the local agent, whether this be the principal's leadership and assistance role, the leadership team's role, or both.

Often, change is not initiated or is not successful because people believe they do not have enough money. Effective school change requires being proactive—grabbing, getting, testing the limits in acquiring needed resources (Huff, Lake, and Schaalman 1982), and taking advantage of potential resources rather than waiting for someone else to provide them. The image comes to mind of school leaders as garage sale junkies, willing and able to browse and find what the school needs in the most unlikely places, thus ensuring support (Bossert et al. 1982).

### Invest in Training and Professional Development

A leadership function that must be satisfied in implementing all efforts to improve schools is that of providing training and professional development (DeBevoise 1984). Skill building and training are part of the process of change. Learning to do something new such as using performance assessment involves initial doubts about one's ability; incremental skill development; some successful experiences; and eventually clarity, meaning, and ownership (Fullan 1985). Effective leaders use formal and informal data to identify needs of the staff for training and development (Little 1982). Fullan's (1985) review of change strategies, Joyce and Showers's (1980) research on staff development, and Sergiovanni's (1990) reports of his experiences in working with leaders in schools suggest investing early in demonstrations and modeling.

Training and development are not accomplished in a one-shot event. Development and growth in new skills is an ongoing process that continues across the initiation and implementation of performance assessment. Many change efforts founder because school systems have simply not provided teachers (and administrators) with the opportunities to acquire the new skills they need; frustration rather than resistance becomes the factor that undermines the planned activities. To demonstrate their commitment, effective leaders participate directly in staff development, taking an active role in planning, conducting, implementing, and evaluating inservice training (Odden 1983; Russell et al. 1985). By participating in staff development with principals and teachers, superintendents demonstrate that they, too, are a part of the community of learners in search of improvement (Murphy, Hallinger, and Peterson 1985).

If school leaders take the responsibility to assist implementers with learning and using new programs, it is quite likely they will need training for this role, whether they are principals, special teachers, central office personnel, or others. Because the goal of training and development is change—in an individual's knowledge, understanding, skills, and behaviors—it is easy to see that school change and teacher development are inextricably linked.

### Assess/Monitor Progress

School improvement efforts, no matter how well-planned, will always encounter problems at all stages. Some are so insignificant that they may not even be perceived as problems. Others are more severe. Louis and Miles (1990) cite the need for continuous monitoring in order to orchestrate the change effort within the school and deal with problems appropriately. Any change effort that is significant in scope—and changing to a comprehensive system of performance assessment *is* significant—or that involves many parts of the school becomes a set of management issues.

Effective leaders constantly search for, confront, and acknowledge serious problems when they first appear and act rapidly to make major adjustments to solve them. They become frequent visitors in the classrooms and take time to discover what is happening. They gather information through formal observation but use informal methods as well, walking the hallways, popping in and out of classrooms, attending grade-level and departmental meetings, and holding spontaneous conversations with individual teachers. They follow these visits with feedback to teachers and plans for improving their use of new—i.e, performance assessment— practices (Rutherford 1985).

In a review of research on principals as instructional leaders, DeBevoise (1984) identified monitoring student and teacher performance as a significant strategy. Superintendents, as well, play an active role in monitoring change and improvement efforts (Pollack et al. 1988). Pollack and colleagues report that superintendents use school and classroom visits to inspect curriculum and instruction and to assess progress in the implementation of new curriculum, instruction, assessment, and other changes. They collect school and classroom products associated with the change. They use student data to monitor the impact of the change on students, as well as to provide input for the superintendents' supervision and evaluation of principals.

Those leading change efforts have a variety of techniques and tools for their use in assessing progress. Formal methods include surveys, observation, and testing; informal methods involve promoting interaction among implementers, between implementers and administrators, and between implementers and other facilitators. Data-collection techniques developed by Hall and colleagues (reported in Hall and Hord 1987), on concerns of teachers and levels of implementation of new practices, are examples of techniques that leaders can use in both formal and informal ways.

### Provide Continuous Assistance

The key to this strategy is the word "continuous." As implementers move from novice to expert, their needs will change. Administrators gather information about these needs through monitoring, then structure assistance to focus appropriately on the needs. For example, if data indicate that some teachers have concerns about how to manage new assessment practices in their classrooms, then the leader can share information, demonstrate or model new approaches, or arrange for teachers to visit classrooms in which other teachers have resolved similar management issues. If monitoring reveals that particular implementers are not putting specific components of the envisioned change into place, leaders can then help the implementers to incorporate the missing pieces into their practice. The leaders can assist by organizing assessment designs, arranging materials, and walking through plans with implementers (Hord et al. 1987).

The continuous assistance strategy involves providing assistance to the implementers in a variety of formats—for example, one-to-one help from peers, administrators, or district resource staff; sharing among implementers; and visits to sites where implementation has occurred. Relatively simple support, such as arranging for teachers' released time to meet *regularly*, can produce important results if administrators address the changing implementation needs of teachers with useful information and tips on how to make the changes work in their classrooms.

The allotment of small, regular blocks of time to foster formal and informal interaction among implementers is essential for continuous assistance. Providing a time and place for such meetings reduces isolation, an important function of a context conducive to change, as noted earlier. Change survives or dies depending on the amount of time and assistance that administrative staff provide—and, one might add, the quality and appropriateness of the assistance. In staff development literature, Joyce and Showers (1980) refer to such assistance as *coaching*. In Bush's (1984) study of the effects of staff development components, coaching accounted for up to 84 percent of the variance in successful transfer of new practices into classrooms.

The ability to provide continuous assistance and be an effective leader for change requires a high tolerance for complexity and ambiguity. Coping with implementation problems is difficult, because staff cannot foresee all the needs. Leaders tend to exhibit a willingness to live with risks as they try various ways to solve persistent issues (Hall and Hord 1987). They also look for positive features, directly and sincerely recognizing and praising the efforts of those responsible (Rutherford 1985). Celebrating progress is an important part of this sixth strategy, an aspect that is often overlooked. Through a dual focus on positive progress and on identification of problems, followed by the necessary corrective action, leaders support the goals and expectations they have established for their schools (Hall and Hord 1987).

It should be obvious that successful implementation of high-quality new programs, processes, or practices will not "just happen." Implementers will need abundant and frequent assistance that the school leadership team can plan for and deliver. It is highly unlikely that one person can assume responsibility for all that a schoolwide or districtwide change such as performance assessment will entail, thus a campus team that shares responsibilities is a practical structure for guiding and supporting school change efforts. Literature on successful change in schools suggests that the action-oriented school team will find the six strategies enumerated above useful as a framework for guiding their efforts in facilitating the implementation of successful change.

## REFERENCES

Anderson, B.L. Personal communication, October 31, 1993.

Block, P. (1987). *The Empowered Manager: Positive Political Skills at Work.* San Francisco: Jossey-Bass.

Blumberg, A., and W. Greenfield. (1980). *The Effective Principal: Perspectives on School Leadership.* Boston: Allyn & Bacon.

Bossert, T., D.C. Dwyer, B. Rowan, and G.V. Lee. (Summer 1982). "The Instructional Management Role of the Principal." *Educational Administration Quarterly* 18, 3: 34-64.

Boyd, V. (1992a). "Creating a Context for Change." *Issues . . . about Change* 2, 2: 1-10. Austin, Tex.: Southwest Educational Development Laboratory.

Boyd, V. (1992b). *School Context: Bridge or Barrier for Change.* Austin, Tex.: Southwest Educational Development Laboratory.

Boyd, V., and S.M. Hord. (1994). *Principals and the New Paradigm: Schools as Learning Communities.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Brandt, R. (1987). "On Leadership and Student Achievement: A Conversation with Richard Andrews." *Educational Leadership* 45, 1: 9-16.

Bridges, W. (1991). *Managing Transitions: Making the Most of Change.* Reading, Mass.: Addison-Wesley.

Bush, R.N. (1984). "Effective Staff Development." *Making Our Schools More Effective: Proceedings of Three State Conferences.* San Francisco: Far West Laboratory for Educational Research and Development.

Chin, R., and K.D. Benne. (1969). "General Strategies for Effecting Changes in Human Systems." In *The Planning of Change,* 2nd ed., edited by W.G. Bennis, K.D. Benne, and R. Chin. New York: Holt, Rinehart, & Winston.

Deal, T.E. (1990). Foreword to *Value-Added Leadership: How to Get Extraordinary Performance in Schools,* by T. Sergiovanni. Orlando, Fla.: Harcourt Brace Jovanovich.

DeBevoise, W. (February 1984). "Synthesis of Research on the Principal as Instructional Leader." *Educational Leadership* 41, 5: 14-20.

Fullan, M.G. (January 1985). "Change Processes and Strategies at the Local Level." *The Elementary School Journal* 84, 3: 391-422.

Fullan, M.G. (1991). *The New Meaning of Educational Change.* New York: Teachers College Press.

Hall, G.E., and S.M. Hord. (1987). *Change in Schools: Facilitating the Process.* Albany: State University of New York Press.

Havelock, R.G. (1971). *Planning for Innovation Through Dissemination and Utilization of Knowledge.* Ann Arbor, Mich.: University of Michigan, Institute for Social Research.

Hord, S.M. (1992). *Facilitative Leadership: The Imperative for Change.* Austin, Tex.: Southwest Educational Development Laboratory.

Hord, S.M., and R. Reynolds-Gibbs. (1993). "The School Board's Leadership Role in Restructuring." *Journal of Texas Public Education* 2, 1: 37-52.

Hord, S.M., W.L. Rutherford, L.L. Huling-Austin, and G.E. Hall. (1987). *Taking Charge of Change.* Alexandria, Va.: Association for Supervision and Curriculum Development.

Huff, S., D. Lake, and M. Schaalman. (1982). *Principal Differences: Excellence in School Leadership and Management.* Boston: McBer.

Jones, B.F., and T.F. Fennimore. (1990). "Videoconference 1: The New Definition of Learning—The First Step to School Reform." *Restructuring to Promote Learning in America's Schools: A Guidebook.* Elmhurst, Ill.: North Central Regional Educational Laboratory.

Joyce, B., and B. Showers. (1980). "Improving Inservice Training: The Messages of Research." *Educational Leadership* 37, 5: 379-385.

Little, J.W. (August-September 1982). "The Effective Principal." *American Education* 18: 38-42.

Louis, K.S., and M.B. Miles. (1990). *Improving the Urban High School, What Works and Why.* New York: Teachers College Press.

MacPhail-Wilson, B., and J. Guth. (1983). "Effectiveness Research and School Administration—Both Sides of the Coin." *NASSP Bulletin* 67, 465: 3-8.

Manasse, A.L. (1984). "Principals as Leaders of High-Performing Systems." *Educational Leadership* 41, 5: 42-46.

Manasse, A.L. (1985). "Improving Conditions for Principal Effectiveness: Policy Implications of Research." *The Elementary School Journal* 85, 3: 439-463.

McCurdy, J. (1983). *The Role of the Principal in Effective Schools* (AASA Critical Issue Report). Sacramento, Calif.: Education News Service.

Méndez-Morse, S. (1992). *Leadership Characteristics that Facilitate School Change.* Austin, Tex.: Southwest Educational Development Laboratory.

Murphy, J. (February 1991). "The 'Maytag Man' of School Reform." *The School Administrator* 48, 2: 32-33.

Murphy, J., P. Hallinger, and K.D. Peterson. (1985). "Supervising and Evaluating Principals: Lessons from Effective Districts." *Educational Leadership* 43, 2: 78-82.

Odden, A. (January 1983). "Research Findings on Effective Schools." ERS *Issuegram.*

Paulu, N. (1988). *Experiences in School Improvement: The Story of 16 American Districts.* Washington, D.C.: U.S. Department of Education, Office of Educational Research and Improvement.

Peterson, K.D. (1984). "Mechanisms of Administrative Control Over Managers in Educational Organizations." *Administrative Science Quarterly* 29: 573-597.

Pollack, S., J. Chrispeels, D. Watson, and S. McCormack. (April 1988). *A Description of District Factors That Assist in the Development of Equity Schools.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Russell, J.S., J.A. Mazzarella, T. White, and S. Maurer. (June 1985). *Linking the Behaviors and Activities of Secondary School Principals to School Effectiveness: A Focus on Effective and Ineffective Behaviors.* Eugene, Oreg.: University of Oregon, Center for Educational Policy and Management.

Rutherford, W.L. (1985). "School Principals as Effective Leaders." *Phi Delta Kappan* 69, 1: 31-34.

Sashkin, M., and J. Egermeier. (1993). *School Change Models and Processes: A Review and Synthesis of Research and Practice.* Washington, D.C.: U.S. Department of Education.

Senge, P.M. (1990). *The Fifth Discipline: The Art and Practice of the Learning Organization.* New York: Doubleday/Currency.

Sergiovanni, T.J. (1990). *Value-Added Leadership: How to Get Extraordinary Performance in Schools.* Orlando, Fla.: Harcourt Brace Jovanovich.

# VIII-9

## Creating a Context for Change

Vicki Boyd
Southwest Educational Development Laboratory

Reprinted with permission from the author. First published in *Issues . . . about Change* 2, 2 (Spring 1992): entire issue.

# CREATING A CONTEXT FOR CHANGE

Improving schools involves change. Change, however, is not an isolated process. It occurs within some context. In the case of school improvement, that context is the school. A school is a complex organism, not just a building with people inside. In order to change the organism, it is necessary to consider the effects of a change on all the parts of the organism. Each part is dependent upon the others and all parts react to changes in any other part. Examining these parts and considering their influence on the change process is important for leaders of school improvement efforts.

Educational research describes several elements of school context that can facilitate the work of leaders of school improvement efforts. What contextual factors influence change? Do these factors also influence success for at-risk students? How might leaders address these factors to support implementation of school improvement for at-risk students?

The impact of the school context on at-risk students is an especially important consideration for those engaged in improving schools. Richardson, Casanova, Placier, and Guilfoyle's study (1989) of at-risk students proposes that at-risk status is derived from an interaction between the characteristics of the child and the nature of the classroom and school. Family background, personal characteristics of the child, the school context, and the social behavior of children interact to create conditions that place children at risk of failing to achieve their academic potential, of dropping out of school, and of having limits placed on their ability to function as productive adults in society.

What is school context? Context is viewed in this article as a broad and inclusive term consisting of two dimensions. The first includes aspects of the school that are not living, but nevertheless affect its inhabitants. The resources available, policies and rules, and size of the school are examples of this dimension of school context, labeled ecology in this paper. The second element is the school's culture. Culture is a term that captures the informal side of schools. It includes attitudes and beliefs, school norms, and relationships, both within the school and between school and community.

As suggested by the Latin *contextere*, "to weave together," the *interrelatedness and interaction* of the school ecology and culture create the context in which school improvement efforts are undertaken. In this paper, the elements of school context are discussed separately. However, it should be emphasized that it is precisely because of the interrelationships and interrelatedness of elements of the school that context is a factor in change (Sarason, 1990).

## Ecology

The school's organization and size, policies, and resources can influence school improvement and at-risk students, because they affect attitudes and relationships among those who inhabit the school.

### The physical surroundings

The physical arrangement and size of schools play an important role in the feelings of teachers and students (Fullan, 1991). Some structures in the school may minimize opportunities for teachers to interact, such as patterns of scheduling, cellular physical layout, and large school size. This, in turn, may promote among educators a decreased sense of self efficacy, an avoidance of controversy, and the often untested assumption that others do not share the same views as the individual (Fowler & Walberg, 1991; Sarason, 1982). Limited contact with new ideas and a decreased capacity for innovation may result from structures that encourage isolation (Fullan, 1991). On the other hand, a sense of community may be developed when schedules and structures are modified to encourage interaction among teachers.

### Formal policies and rules

In addition to the physical plan of the school, policies at federal, state, and district levels can present barriers or bridges to school improvement efforts and the success of at-risk students. In fact, it has been argued that departure from district or state policies and regulations is often required to achieve substantial positive change in instruction (Levine, 1991). Examples of policies likely to assist school improvement effort include those that provide greater autonomy at the building level, foster collaboration among schools and teachers, and provide effective channels for communication and staff development (Carnegie Foundation, 1988; Chubb & Moe, 1990).

### Resources

The availability of local school resources has great influence on school improvement implementation. Locating and allocating both time and money for the change project are major responsibilities of leaders (Louis & Miles, 1990). Finding time to engage in implementation activities is often cited as a barrier to implementation. Researchers have found that, indeed, many change efforts fail simply because not enough time was invested in them (Sarason, 1982; Simpson, 1990). In addition, the lack of funding can limit the type of improvement considered, limit the materials available, and result in the inability to address problems until the next fiscal year (Louis & Miles, 1990; Pink, 1990)

## The School Culture

The culture of the school can assist improvement efforts for at-risk students, or act as a barrier to positive change (Deal & Kennedy, 1982; Sarason, 1982). A summary of the general knowledge base regarding school culture by Patterson, Purkey, and Parker (1986) provides insight for leaders in examining their school culture. They found:

- School culture affects the behavior and achievement of students.
- School culture is created and can be shaped by people in the school.
- No two school cultures are exactly alike.
- The culture bonds the individuals in a school together when it provides focus and a clear purpose for the school.
- Culture can be an obstacle to educational success; it can be oppressive and discriminatory for various subgroups.
- Lasting change in fundamental areas such as teaching practice or decision-making structures requires understanding and altering the school culture. This is a slow process.

### Attitudes

The attitudes of teachers regarding schooling, students, and change have an impact on the behavior of teachers toward students, especially those at risk. These same attitudes influence teachers' behavior in implementing or resisting school improvement as well. The influence of teacher attitudes is particularly important due to the interplay between the characteristics of the student and the context of the school that defines a student as at risk.

Identifying and confronting beliefs among school staffs that prohibit students from achieving their potential is a vital component of school improvement efforts. Commonly accepted myths become barriers when they limit students' access to quality education. Some of these myths, identified in a study by the Massachusetts Institute of Technology (1990), include:

- Learning is due to innate abilities, and minorities are simply less capable of educational excellence than whites.
- The situation is hopeless.
- Education is an expense and not an investment.
- Equity and excellence in education are in conflict.
- Minorities don't care about education.
- Educational success or failure is within the complete control of each individual.

Negative attitudes toward change can also hamper efforts to improve schools. A "system paranoia" that is reflected in statements such as "they won't let me do it," or "I knew things hadn't changed," or "there they go again" creates barriers to change (Goldman & O'Shea, 1990). Teacher concerns about how change will affect them personally and the practicality of change are areas to be addressed by leaders. Examining change efforts of the past is important because, if these past experiences were unsuccessful, teachers may be cynical or apathetic toward the new school improvement effort (Corbett, Dawson, & Firestone, 1984; Fullan, 1991). Taking time to ensure that the reasons for the change, the practicality of this program for the specific problem being addressed, and the philosophical basis for the effort are well understood by everyone involved will enhance the likelihood of lasting implementation (Boyd, 1992).

Students' attitudes toward change, though examined little, are significant for the success of school improvement. Fullan (1991), who notes that few researchers have actually *asked* students what they think, proposes four images of student attitudes toward change: indifference, confusion, a temporary escape from boredom, or heightened interest and engagement with learning. If the school improvement effort creates conditions where students become actively engaged, it is more likely that increased student success will occur as a result of the effort (Fullan, 1991).

In addition, the community's attitudes toward the school and the school improvement efforts affect implementation. External groups exert pressure on schools requiring schools to satisfy what their constituents believe is proper for schools (Cuban, 1990). If the community provides ongoing encouragement, support and resources, change efforts will more likely enjoy success (Goodlad, 1984). Developing this support by the community is an important task of school leaders.

### Cultural norms

Just as the attitudes and beliefs of persons both inside and outside the school building may facilitate or impede change, the norms, or informal rules that govern behavior, exert influence on change efforts. Researchers have found particular cultural norms that facilitate school improvement: a norm of continuous critical inquiry, a norm of continuous improvement, a widely shared sense of purpose, and a norm of involvement in making decisions (Barth, 1991; Louis & Miles, 1990; Saphier and King, 1985).

A norm of continuous critical inquiry may indicate that school personnel are aware of the school's strengths and weaknesses and are open to dealing with its imperfections. Programs that are successful for at-risk students do not suppress criticism or emphasize conformity (Fine, 1991). Introspection and reflection about the beliefs, actions, and underlying assumptions that regulate teaching and administration are important to successful change efforts (Dreyfuss, Cistone, and Divita, 1992).

A close companion of critical inquiry is the idea that continuous improvement is expected. Experimentation and staff development lead to improved practice and an increased capacity for change through the development of knowledge, skills, and understanding. The expectation that staff development will occur is a norm that promotes change efforts. As Miles and Louis (1990) point out, "knowing that X is a workable action you want to take does not mean knowing how to deliver X. The paths to educational improvement are strewn with examples of behavior that no one knew how to deliver" (p.58).

The need for a widely shared sense of purpose or vision has been described in other *Issues...about Change*, as has the idea that broad-based involvement in decision-making, especially decisions regarding teaching and curricular issues, is important to change efforts. It bears repeating, however, that improved student outcomes that are clearly articulated, relevant, and visualized in action images must be the focus of whatever change schools undertake (Miles & Louis, 1990).

When the existing norms of the school encourage continuous introspection, continuous improvement, and involvement in decision making, change is encouraged. The development of such norms is an important aspect of leadership for change.

### Relationships

Just as individuals' attitudes and beliefs affect change and the norms of the school, relationships among individuals and groups are part of the school culture that can either facilitate or impede change. In turn, these relationships are affected by cultural norms.

**Teacher peer relationships.** A norm of collegial relationships is often cited as a component of successful change implementation. A collaborative work culture can reduce professional isolation and contribute to higher morale, enthusiasm, and willingness to accept new ideas (Barth, 1991). Leaders of school improvement efforts need to encourage and support collegial relationships between principal and teachers and among teachers. Collegial relationships facilitate change because peer group consensus and interaction are important to social learning (Goodlad, 1984), and change has been characterized as a learning process.

**Student-teacher relationships.** In the traditional, isolated professional culture, the teacher develops mental models of schools and students based on the ways students respond to what the teacher is doing. Indeed, as Rosenholtz (1987) found, teachers rely primarily on students to provide meaning for their work. This reliance may encourage gradual narrowing of the teacher's sense of purpose, decreased teacher expectations for students, and increased weight given to practices that enable the teacher to manage classroom behavior successfully. Rarely, in an isolated professional culture, will teachers' assumptions, norms, values and beliefs be challenged by significantly more ambitious visions of what is possible (Leithwood & Jantzi, 1990).

Change efforts will have limited success unless students are actively engaged. The attitudes of students, the influence of cultural norms on their behavior, and the relationships they have with teachers, other students and the school itself affects school improvement efforts (Boyd, 1992).

High school students often develop a "live and let live" relationship with teachers that presents a barrier to change by protecting the status quo (Fullan, 1991). Research has shown that a cycle occurs where teachers' expectations affect students' commitment, which then affects teachers' expectations. (Firestone & Rosenblum, 1988).

At-risk students have the same needs as all adolescents for group membership, positive relationships with adults, acquisition of knowledge and skills, and a sense of competence (Wehlage, Rutter, Smith, Lesko, & Fernandez, 1989). Peer leaders and student heroes affect the scholastic tone of the school and enhance student achievement (Deal, 1985). A school context that forms a sense of community is necessary to promote student growth (Smey-Richman, 1991).

**Relationships with parents and community**. Transforming schools entails a fundamental change in traditional beliefs held by parents, local politicians, residents, administrators, staff, and students (Deal, 1990). Goals, standards, and practices of the school must be consistent with, and developed in partnership with, parents and community members. Community agreement on desired schooling outcomes is essential for effective schooling to take place, as is community participation in reviewing progress toward these desired outcomes (Weber, 1986 cited in Conklin & Olson, 1988). Community support of the school itself and for the change effort is vital. Parents and community members must be active partners and allies, not adversaries.

### Leadership and Context

Purkey and Smith (1983), in their review of the effective schools literature, discuss the idea that schools are "loosely coupled" systems in that what goes on in individual classrooms is largely independent of mandates and supervision. As noted earlier in this article, most classrooms are isolated, and the physical structure of schools contributes to this isolation. Sergiovanni (1987) proposes that what is needed in schools is a combination of a tightly coupled culture and loosely coupled structure. A combination of tightly coupled culture, "around clear and explicit themes representing the core of the school's culture," and of loosely coupled structure, "so that people can pursue these themes in ways that make sense to them, may well be a key reason why [successful] schools are so successful" (Sergiovanni, 1987, p. 62).

Creating a context that contains a tightly coupled culture may require leaders to view schools from a different perspective.

> [The cultural perspective] assumes that consensus is more powerful than overt control, *without ignoring the need for leadership*. Building staff agreement on specified norms and goals becomes the focus of any school improvement strategy. It is important to note, however, that staffs could agree on educationally unsound ideas of practices; consensus could also act as a buffer and prevent critical examination of the school or a proposed change. While staff agreement is important, we do not mean that it should be uncritical (Purkey & Smith, 1983, p. 442).

A focus on improved educational experiences and success for all students should guide the efforts.

The stories and suggestions that follow are drawn from the experiences of particular schools in implementing change and implementation research. The reader is reminded, however, that due to the interrelational and interdependent nature of the elements of school context, specific activities that worked in one school may not work as well in another. As Leithwood (1992) observes, "we are learning that schools are complex systems made up of parts with greater interdependencies than we earlier believed" (p. 9). This interdependence is the challenge provided by context to those leading school improvement efforts.

## Shaping the ecology

The principal needs to identify structural components of the school that influence change efforts. As the person typically responsible for scheduling and physical arrangements in the school, the principal can use these tools to reduce isolation and alienation of students and teachers. For example, principals might schedule common planning periods for teachers to share information and interact. Through allocation of resources, principals can give preference to teachers actively engaged in collaborative practices or to field trips that involve two or more teachers working together on a project (Smith & Scott, 1990). This sets the stage for the development of collegial relationships.

Smith and Scott (1990) suggest several ways in which principals can use faculty meetings as a vehicle for encouraging collaboration: Allow faculty to suggest agenda items; engage faculty in chairing the meetings on a rotating basis; hold the meeting in various locations to allow each teacher the opportunity to serve as host; set aside time at faculty meeting for teachers to describe inservice programs they have attended and to highlight teacher achievements.

Simpson (1990) tells the story of the principal of a school that had been engaged in school-wide improvement for nine years. The principal began by inviting anyone interested to join her for lunch one day a week to discuss improvement efforts. Outside consultants sometimes presented ideas to the group and even arranged a visit to another school. Nine years later, these lunches had become a "protected ritual" and an important time together for planning. These lunch meetings led to such changes as thematic curriculum, cross-age teaching, and exhibits of student work. The principal cited rituals such as the lunch meeting as vehicles for change in the school.

## Nurturing and supporting a positive, widely internalized school culture

Deal and Peterson (1990) point out that principals who are trying to understand the culture of their school need to ask three questions:

- What is the culture of the school now — its history, values, traditions, assumptions, beliefs, and ways?
- Where it matches my conception of a "good" school, what can I do to strengthen existing patterns?
- Where I see a need for new direction, what can be done to change or reshape the culture? (p. 16)

In order to answer the first question, the principal must observe what is happening in the school. One possible framework for analyzing the culture is suggested by the work of Deal (1985). By examining the shared values and beliefs, heroes and heroines, rituals and ceremonies, stories told, and the informal network of cultural players, principals can begin to understand the culture of the school.

- Shared values and beliefs may be uncovered by looking at written documents such as the school's mission statement or any publications written by the school that contain slogans or symbols.
- The heroes or heroines are those individuals such as former and current teachers, principals or other staff whose heroic deeds or qualities serve as current or past role models.
- Rituals and ceremonies (such as where and when meetings are held; how new employees are oriented; how accomplishments are recognized; who socializes with whom, when, and where) provide insight about what is considered valuable.
- Stories that are told in the school about heroic deeds or the success of certain teaching practices carry messages

regarding cultural values.

- The informal network that controls communication patterns and channels may be analyzed by considering who talks and writes to whom, when, why, and the response they get; who in the school serves as a gossip, spy, or soothsayer.

Once the principal has investigated these aspects of the school's culture, an examination of ways to strengthen or reshape existing cultural patterns may be undertaken. According to Schein (1985), the primary mechanisms principals use to strengthen existing cultural elements that fit with the shared purpose of the school and reshape cultural elements that conflict with the vision are:

- what they pay attention to,
- their reactions to critical incidents and crises,
- their role modeling, teaching, and coaching, and
- criteria they use to recruit, reward, and censure employees. (p. 225)

**Paying attention.** Researchers (e.g. Deal & Peterson, 1990; Leithwood & Jantzi, 1991) who have studied principals who shape the school culture have found some common behaviors and tactics used by these principals. Those who shape culture have a clear and focused sense of mission and values and develop a vision of what the school should be. They strengthen a widely internalized culture through such activities as clarifying and prioritizing a set of shared goals. For example, by asking questions regarding how implementation is proceeding and what problems are occurring, principals demonstrate the importance of the school improvement effort. The principal is also communicating an "expectation to be informed about classroom practices" (Staessens, 1991, p. 4). Paying attention to how the shared goals are being played out in classrooms is demonstrated in this way.

**Modeling.** The principal is extremely important to the school's norms and relationships. The principals studied by researchers such as Deal and Peterson (1990), Leithwood and Jantzi (1991), and others actively model their values in a highly visible manner through consistent daily routines and concrete actions. Principals demonstrate what is important in the school through activities such as allocating resources, planning and scheduling time for collaboration, and involving others in decision-making. By modeling a high degree of interaction with the staff, the principal encourages a high degree of cooperation among the staff (Smith & Scott, 1990). Norms such as "one should work hard" are created by principals who model this norm. "He is the first to be at school, and the last to leave; The school is his hobby" are teachers' remarks about one principal who provided such modeling (Staessens, 1991, p.11).

**Teaching and coaching.** Staessens (1991) found that principals who nurture and support a culture conducive to change were well read and well informed. By making sure that answers and help are provided for problems encountered by teachers, principals encourage a norm that "the school is a place where teachers can learn something, and that one can become a better teacher by bringing up something in a professional manner" (Staessens, 1991, p. 11). Principals foster teacher learning by such activities as providing resources for staff development, attending in-service training with the teachers, and sharing information from conferences with teachers. Stories that illustrate what principals value in the school are spread to become "legends" that will strengthen the culture even further. Frequent and direct communication about norms, values, and beliefs further strengthens a positive school culture.

**Reacting to critical events.** Stories about reactions to critical events are a tool used to strengthen the culture. For example, one principal tells the story of "cart teachers" known for their "can do" attitude. Due to a shortage of space, these teachers kept materials on a cart and moved from room to room, exemplifying the school's focus on solving problems creatively (Blendinger & Jones, 1989, p. 24). When conflict occurs, principals who nurture and support a positive culture are willing to face conflict rather than avoid it. They use structures such as faculty meetings and common planning periods to reduce teacher isolation.These principals use conflict as a vehicle to resolve disputes and build unity.

**Selecting, rewarding, and censuring staff.** Finally, each of the principals in the studies cited above used to some degree traditions, ceremonies, rituals, and symbols that expressed and reinforced the school culture he or she was striving to construct. These leaders used symbols and rituals (such as staff meetings and assemblies where the work of staff and students was recognized, private notes that expressed appreciation for special efforts, and encouragement to share experiences with colleagues) to express cultural values. Teacher commitment to shared goals is stimulated by principals, sometimes through such forceful means as giving teachers the option to transfer elsewhere with the principal's assistance if they do not want to stay and devote themselves to the goals of the school. Teacher evaluation has also been found to contribute significantly to teachers' commitment to school goals (Leithwood & Jantzi, 1991).

Although this paper highlights the role of the principal as leader in the school, Patterson, Purkey, and Parker (1986) point out that the "principal is not, in fact, the only person who can provide leadership, especially leadership for school improvement" (p. 103). They go on to note that school context is likely to "strongly affect what a principal should (and can) do in any given situation" (p.104). Because of the influence of school context on school improvement efforts, leaders must understand that schools are complex organisms. The fact that the leader is also part of this organism increases the need to understand and learn how to work with the elements of school context if school improvement is to succeed.

Successful implementation of school improvement for at-risk students occurs where leaders have "vision, the ability to generate solutions, and the foresight to provide security to teachers as they try to accommodate to new problems" (Davis, 1989, p.6). Once improvement efforts are underway, principals must be aware of the changing demands on their leadership. According to one principal, "It's all the things that can go wrong after a program is adopted that burn people out. People like me must see to it that teachers who want to make improvements are supported" (Avery, 1986, p. 4).

It is important that leaders of school improvement efforts pay attention to the school context. Understanding that elements of context affect change is the first step toward dealing with the influence of these elements. Such actions as the modification of schedules and allocation of resources by leaders can support school improvement. Structures and rituals can be initiated that will provide vehicles for change. By examining cultural indicators and working to strengthen those elements of the culture that fit the school improvement effort, leaders can create a context that supports change.

*Issues...about Change* is published and produced quarterly by the Leadership for Change Project of the Southwest Educational Development Laboratory. This issue was written by Vicki Boyd, Research Associate, Services for School Improvement, SEDL.

BEST COPY AVAILABLE

## References

*An imperiled generation: Saving urban schools.* (1988). A special report of the Carnegie Foundation for the Advancement of Teaching. Princeton, NJ: Princeton University Press.

Avery, R. (1986). The urban school principal: The rocky road to instructional leader. *Carnegie Quarterly, 31*(1), pp. 3-7.

Barth, R.S. (1991). Restructuring schools: some questions for teachers and principals. *Phi Delta Kappan. 73* (2), pp. 123-128.

Blendinger, J. & Jones, L.T. (1989). Start with culture to improve your schools. *The School Administrator*, pp. 22-25.

Boyd, V. (1992). *School context: Bridge or barrier to change?* Austin, TX: Southwest Educational Development Laboratory.

Chubb, J.E. & Moe, T.M. (1990). *Politics, markets, and America's schools.* Washington, D.C.: The Brookings Institution.

Clune, W.H. (July, 1991). *Systemic educational policy.* Madison, Wisconsin: Wisconsin Center for Educational Policy.

Conklin, N.F. & Olson, T.A. (November 1988). *Toward more effective education for poor, minority students in rural areas: what the research suggests.* Portland, OR: Northwest Regional Educational Laboratory.

Corbett, H., Dawson, J., & Firestone, W. (1984). *School context and school change: Implications for effective planning.* Philadelphia, PA: Research for Better Schools.

Cuban, L. (1990). Reforming again, again, and again. *Educational Researcher, 19* (1), pp. 10-21.

Davis, J. (Winter 1988-89). Culture and change in the small school. *The Rural Educator, 10* (2), pp. 4-7.

Deal, T. (1985). The symbolism of effective schools. *The Elementary School Journal, 85* (5), pp. 601-620.

Deal, T. (1990). Reframing Reform. *Educational Leadership.,47* (8), p. 6-12.

Deal, T. & Kennedy, A. (1982). *Corporate cultures.* Reading, MA: Addison-Wesley Publishing Co, Inc.

Deal, T. & Peterson, K. (1990). *The principal's role in shaping school culture.* Washington, D.C.: U.S. Department of Education.

Dreyfuss, G.O., Cistone, P.J., & Divita, C. (1992). Restructuring in a large district: Dade County, Florida. In C.D. Glickman (Ed.), *Supervision in transition* (pp.77 - 96). Alexandria, VA: Association for Supervision and Curriculum Development.

*Education that works: An action plan for the education of minorities.* (1990). Quality Education for Minorities Project. Cambridge, Mass: Massachusetts Institute of Technology.

Fine, M. (1991). *Framing dropouts: Notes on the politics of an urban public high school.* Albany, NY: State University of New York Press.

Firestone, W.A. & Rosenblum, S. (1988). Building commitment in urban high schools. *Educational Evaluation and Policy Analysis, 10* (4), pp. 285-299.

Fowler, W.J. & Walberg, H.J. (1991). School size, characteristics, and outcomes. *Educational Evaluation and Policy Analysis, 13*: (2), pp. 189-202.

Fullan, M.G. (1991). *The new meaning of educational change, 2nd Edition.* New York: Teachers College Press.

Grimmett, P.P., Rostad, O.P., & Ford, B. (1992). The transformation of supervision in C. D.Glickman (Ed.), *Supervision in transition.* (pp. 184-202). Alexandria, VA: Association for Supervision and Curriculum Development.

Goldman, C. & O'Shea, C. (1990). A culture for change. *Educational Leadership, 47*(8), pp. 41-43.

Goodlad, J. (1984). A place called school - Prospects for the future. New York: McGraw-Hill.

Leithwood, K.A. (1992). The move toward transformational leadership. *Educational Leadership.,49* (5), pp. 8-10.

Leithwood, K.A. & Jantzi, D. (1990). Transformational leadership: How principals can help reform school cultures. *School Effectiveness and School Improvement*, 1(4), pp. 249-280.

Levine, D.U. (1991). Creating effective schools: Findings and implications from research and practice. *Phi Delta Kappan, 72* (5), pp. 389-393.

Louis, K.S. & Miles, M.B. (1990). *Improving the urban high school*. New York: Teachers College Press.

Miles, M.B. & Louis, K.S. (1990). Mustering the will and skill for change. *Educational Leadership, 47* (8), pp. 57-61.

Patterson, J.L., Purkey, S.C., & Parker, J.V. (1986). *Productive school systems for a nonrational world*. Alexandria, VA: Association for Supervision and Curriculum Development.

Pink, W.T. (1990). Staff development for urban school improvement: Lessons learned from two case studies. *School Effectiveness and School Improvement, 1*(1), pp. 41-60.

Purkey, S.C. & Smith, M.S. (1983). Effective schools: A review. *The Elementary School Journal, 83* (4), pp. 427-452.

Richardson, V., Casanova,U., Placier, P., & Guilfoyle, K. (1989). *School children at-risk*. New York: The Falmer Press.

Rosenholtz, S. (1987). School success and the organizational conditions of teaching. In J.J. Lane & H.J. Walberg (Eds.), *Effective school leadership*. (pp. 187-218). Berkeley, CA: McCutchan Publishing.

Saphier, J. & King, M. (1985). Good seeds grow in strong cultures. *Educational Leadership, 42* (6), pp. 67-74.

Sarason, S.B. (1982). *Culture of the school and the problem of change, 2nd Edition*. Boston, MA: Allyn & Bacon.

Sarason, S.B. (1990). *The predictable failure of educational reform - Can we change course before it's too late?* San Francisco: Jossey Bass.

Schein, E.H. (1985). *Organizational culture and leadership*. San Francisco: Jossey Bass.

Sergiovanni, T.J. (1987). *The principalship*. Boston: Allyn and Bacon, Inc.

Shields, P.M. (1990). A review of research on school and community influences on effective curriculum and instruction. In M.S. Knapp and P.M. Shields (Eds.), *Better schooling for the children of poverty*. (pp. XIII-3-XIII-10). Washington, D. C.: U.S. Department of Education.

Simpson, G. (1990). Keeping it alive: Elements of school culture that sustain innovation. *Educational Leadership, 47* (8), pp. 34-37.

Smey-Richman, B. (1991). *School climate and restructuring for low-achieving students*. Philadelphia, PA: Research for Better Schools.

Smith, S.C. & Scott, J.J. (1990). *The collaborative school: A work environment for effective instruction*. University of Oregon, Eugene: ERIC Clearinghouse of Educational Management.

Staessens, K. (1991). *The professional culture of innovating primary schools: Nine case studies*. A paper presented at the annual meeting of the American Educational Research Association, Chicago, 1991.

Wehlage, G.G., Rutter, R.A., Smith, G.A., Lesko, N., & Fernandez, R.R. (1989). *Reducing the risk: Schools as communities of support*. New York: The Falmer Press.

**SEDL | OERI**

This publication is based on work sponsored wholly, or in part, by the Office of Educational Research & Improvement, U.S. Department of Education, under Contract Number RP91002003. The contents of this publication do not necessarily reflect the views of OERI, the Department, or any other agency of the U. S. Government.

This publication may be reproduced and copies distributed to others. Please acknowledge SEDL as the source in all copies.

# VIII-10

# Realizing School Improvement Through Undersanding the Change Process

SHIRLEY M. HORD
SOUTHWEST EDUCATIONAL DEVELOPMENT LABORATORY

Reprinted with permission from the author. First published in *Issues . . . about Change* 1, 1 (Winter 1990): entire issue.

## Realizing School Improvement Through Understanding the Change Process

Federal legislation, state regulations, district mandates – in addition to local school and community interests – all have demanded change, improvement, and new results from schools. These voices have vigorously exhorted school administrators and teachers to respond to the challenges of producing a more technologically prepared citizenry and to the needs of various student populations designated as at-risk. While schools have responded with effective schools projects, school improvement programs, and any number of efforts to change curriculum and instructional practices, little real or widespread change seems to have occurred. Schools, classroom settings, teaching, and instructional strategies are remarkably the same as they were years ago (Brandt, 1991). Cuban (1988, p. 89) observes that "innovation after innovation has been introduced into school after school . . . disappearing without a fingerprint." Why has so little change been realized? Why have so few new results been obtained?

Nearly everyone is familiar with the number of highly promising new programs or innovations that have paraded through schools. Their recurring introductions were matched all too frequently by teachers' antipathy and resignation. "Don't put too much energy into this one," they would say, "this, too, will go away." This is not to fault teachers, who typically received little support, guidance, or assistance in implementing the new programs. Rather, the change process was not given sufficient attention.

The regular demise of this multiplicity of programs led to the judgement that the innovations themselves were not effective, because they produced no appreciable student outcomes. In most cases, in fact, the programs were never incorporated into the day to day operations of classrooms and could be found in bookshelves gathering dust. As Fullan (1982, p. 22) suggests, "Reforms . . . may fail to make a difference . . . because the plans and resources necessary to accomplish implementation are not adequate to the task." A simple way to state this lack of events is: no implementation, no change, no improved results.

The good news is that we understand a great deal about the process of change and how to incorporate new programs into classroom life. *Acting* on that understanding can provide a significant means for moving schools toward desired improvement. One of the imperatives for acting immediately to address school change and improvement is the status of many children in our society, and their widespread lack of success in schools. These at-risk students now constitute thirty percent of the school population. They include students who will drop out before high school graduation, or who will graduate ill-prepared for meaningful work or post-secondary education (Knapp & Schield, 1990). "We are programming our children and our nation to fail," states Charles Gershenson, director of statistical analysis for the Center for the Study of Social Policy (Goldman, March 31, 1991). Policymakers in both the education and economic sectors call for action now. Fortunately a number of innovative programs and strategies, tested for their potential to increase the educational success of at-risk students, are available (Orr, 1987). However, no innovation, however effective, can succeed unless schools accommodate and address the process of change.

Two important questions, each focused on a dimension of the change process, can inform implementation activities so that school improvement can be realized: What exactly is the innovation to be introduced? What happens to the individuals who will implement the innovation into their work settings and practices? A discussion of these two questions is the major focus of this paper.

### The Innovation

In the 1970s a Rand Corporation study of federal programs supporting educational change revealed interesting findings regarding the implementation of these programs by school staff. One finding, labeled "mutual adaptation," pointed to the researchers' insights about the substantively different characteristics that the programs took on as they were implemented by teachers (Greenwood, Mann, & McLaughlin, 1975). In the same period, this phenomenon was reported also by Hall and Loucks (cited in Hall & Hord, 1987), who concluded that any one innovation usually appeared to have multiple *configurations* or *operational patterns* resulting from variations in teacher selection and use of the components of the innovations.

Such differences in the innovation's implementation in classrooms, or Innovation Configurations (Hall & Hord, 1987), were attributed to several factors, including: how the new program was introduced to teachers (and other intended users) in terms of clarity of description of the program, how complex the program was to incorporate into classroom practice, whether expectations for use of the program were established, and whether assistance was provided to potential users in their implementation efforts. Through discussions with many teachers and other educational staff involved in change, researchers have learned that practitioners typically lack clarity about what the innovation *is* at its introduction and subsequently what it should look like *during* implementation. Such a lack of definition leaves intended users with little understanding of the demands or requirements of the innovation and with impoverished mental images of what it should look like if it has been implemented in a high quality way. This problem has been found not only with innovations such as curriculum programs, instructional processes, strategies, and other new practices, but also with large-scale changes (such as restructuring) that seek to alter the fundamental ways schools operate (Van den Berg, 1981).

The effects of such lack of definition, or vision, have been demonstrated very clearly with, for instance, hands-on inquiry science programs. Inquiry science is predicated on students themselves interacting with materials, learning the skills of scientists through their own science investigation experiences, and constructing their own knowledge. There have been excellent examples of implementation of this model by teachers; however, what is called "inquiry science" can be found in classrooms where the teacher is the sole person manipulating the materials and the experiences, while students look on passively – not the intended operational form of the program at all. It is easy to understand in this case that what students learn from the *actual* implementation is far different from what was hoped to be gained from the *intended* implementation. Under such circumstances, hoped-for outcomes cannot be realized (James & Hord, 1988).

Both a tool and techniques for its use have been developed that enable the identification, articulation, and description of the *components* of an innovation. The framework of Innovation Configurations is this tool (Hall & Hord, 1987); it also includes a process whereby certain *variations* in implementing the components in classrooms are deemed "ideal," "acceptable," or "unacceptable," thus identifying areas of latitude in implementation for teachers and others. The message here is not that all teachers should necessarily be expected to implement all programs in an identical fashion; fidelity is not the theme. The theme is that intended users should be privileged to understand what the new program is all about, what it is expected to produce, and the various ways in which its implementation may produce the desired results. Until such issues as these are addressed, there can be little expectation that the desired change, and school improvement, will be reached through implementors' efforts.

### The Implementors

Rogers (1971) reviewed studies of agricultural extension agents and farmers, and of doctors integrating penicillin and other medications into their practice, and found that participants in the studies differed in their readiness to accept change. Some adopted the change quickly; others took a much longer time. Rogers identified five categories of people relative to their rate of adoption of innovations: innovators, those persons eager to try new ideas, were open to change and willing to take risks, and were frequently perceived as a bit naive or crazy; leaders, open to change, were more thoughtful about getting involved, and were sought for their advice and opinions; the early majority, the people who were cautious and deliberate about deciding to adopt an innovation, tended to be followers, not leaders; the late majority, those skeptical of adopting new ideas, were characterized as "set in their ways"; and resisters, suspicious and generally opposed to new ideas, were usually low in influence and often isolated from the mainstream.

Shortly after Rogers reported his typology, Hall and colleagues, working in teacher education, identified categories of teachers who were in the process of becoming a teacher, or who were adopting and implementing new teaching strategies, or curricula, etc. (cited in Hall & Hord, 1987; and Hord, Rutherford, Huling-Austin, & Hall, 1987). The categories were based on teachers' reactions or concerns as they experienced the adoption and implementation processes. In this articulation, seven kinds of concern that individuals have about an innovation were identified. Each of the stages was distinguished from the other, but they were not mutually exclusive. That is, individuals could have concerns at all stages at any particular time, but the intensity of specific stages varied with any individual as implementation progressed.

The seven stages have been defined by the expressions of concern that individuals make at the various stages. When the individual is not related to or influenced by the innovation in any way, concerns are typically characterized as "not being concerned about it," and this is stage 0. When a new program or innovation appears to have the potential to become part of an individual's life, then stage 1, or informational concerns, become uppermost. These concerns relate to an interest in acquiring information about the innovation. For instance, when a secondary school administrator telephones the office of the secondary school principals' association and makes inquiries about school-based management and how it works, this exemplifies informational concerns about the innovation of school-based management.

As informational concerns develop, stage 2 personal concerns become highly predictable. The individual wonders how the new program will affect him/her and whether he/she may have the ability to work appropriately with the innovation. The introduction of shared decision-making, for example, generally stimulates personal concerns for both school-based administrators and teachers as they contemplate how their current roles and relationships may change as a result of the new decision-making process. Administrators may wonder about the status of their positional authority in the new procedures, and teachers may question their capacity to be involved in school-wide matters.

As individuals begin implementation, stage 3 management concerns become intense. It is easy to understand that teachers have high management concerns while they are learning to cope with the demands of new programs and practices. This is especially true during implementation of an innovation such as cooperative learning, where restructured arrangements of students and teachers are expected. Managing a roomful of student committees, each structuring their own learning experiences within their membership, produces teachers' concerns about controlling time, noise-level, activity, and student behaviors. At this point, time to understand the requirements and logistics of putting new practices into place become factors in high need of attention. "Mastering" management and its attendant concerns will require substantial time.

Hall and associates discovered that, once the intensity of concerns at stages 1, 2, and 3 are reduced, then there is the possibility that individuals will increase their stage 4 consequence concerns. Consequence concerns focus on the effects of the innovation on students, and interest by the individual in making the innovation and its use more effective for students (or whomever the "client" is). For instance, teachers who have been developing students' critical thinking skills as an isolated discipline, may, after they feel comfortable with the strategies involved, determine that critical thinking would be more effective for their students if it were integrated with academic subject areas. Considering this kind of action would reflect consequence concerns.

Many individuals will never reach either intense stage 5 collaboration concerns, which relate to collaborating with others to increase the outcomes of the innovations, or stage 6 refocusing concerns, which focus on major ways to enhance or change the program to further increase effectiveness. Frequently such collaboration and refocusing are forestalled by new demands and innovations "coming down the pike" to schools.

The Stages of Concern About an Innovation concept and its assessment procedures have been productively utilized in a large number of studies of implementation and change at all levels of education — kindergarten through college — and for a vast array of innovations. Such use builds confidence that these stages, or categories, as descriptions of people progressing through the change process, are reasonably accurate. The research suggests that people differ in their approach and response to and movement through change, and that they may require differential support and assistance for success. The Stages of Concern approach and procedures can be used to help gauge and understand individuals and their needs as they experience implementation and its related concerns.

## So What?

It seems reasonable to suggest that "business as usual" will produce results as usual. Thus, if different results are desired, then "business" in the school and classroom will, of necessity, have to be different. Changes in the school staff's knowledge, understanding, skills, and behaviors will be required. This article argues that an innovation's demands for new skills and behaviors, and the individual's needs while implementing the new practices, will require interventions that give attention and support as the change process unfolds. Certainly, the organizational context in which implementation takes place is another critical factor for consideration, and it is the subject of another article in this series on leadership for change.

Although the issues of implementation raised here apply generically to any change effort, the Leadership for Change Project in the Southwest Educational Development Laboratory that is developing this series of articles, is focusing attention on the needs of at-risk students and the role that leadership at all educational levels can exercise in these students' behalf. To reach at-risk students with potentially powerful programs and strategies that increase student success, those programs must be implemented well in daily school and classroom practice. Supporting implementation is a function of leadership.

To conclude, when educational leaders *understand and acknowledge* that the change process itself is a factor to be *accommodated* in their school improvement efforts, when they consider the *requirements of the changes or innovations* that are introduced and the *needs of all individuals* who will be implementing the innovations, and when they *develop plans* that take these factors into account, then they will be providing leadership that guides, manages, and supports change. Only then will the likelihood of school improvement be realized.

## References

Brandt, R. (1991). Coping with change. *Educational Leadership, 48*(7), 3.

Cuban, L. (1988). Constancy and change in schools (1980s to the present). In P.W. Jackson (Ed.), *Contributing to educational change*. Berkely, CA: McCutchan Publishing.

Fullan, M. (1982). *The meaning of educational change*. New York: Teachers' College Press.

Goldman, H. (1991, March). Black children get left to fall behind at the starting line. *Austin American Statesman*, p. H1.

Greenwood, P.W., Mann, D., & McLaughlin, M.W. (1975). *Federal programs supporting educational change, vol. III: The process of change*. Santa Monica, CA: Rand Corporation.

Hall, G.E., & Hord, S.M. (1987). *Change in schools: Facilitating the process*. Albany, NY: State University of New York Press.

Hord, S.M., Rutherford, W.L., Huling-Austin, L., & Hall, G.E. (1987). *Taking charge of change*. Alexandria, VA: Association for Supervision and Curriculum Development.

James, R. K., & Hord, S. M. (1988). Implementing elementary school science programs. *School Science and Mathematics, 88*(4), 315-334.

Knapp, M. S., & Shields, P. M. (1990). Reconceiving academic instruction for the children of poverty. *Phi Delta Kappan, 71*(10), 753-758.

Orr, M. T. (1987). *Keeping students in school*. San Francisco: Jossey-Bass.

Rogers, E. (1971). *Diffusion of innovations*. New York: The Free Press.

Van Den Berg, R. (1981). *Large scale strategies for supporting complex innovations in participating schools*. 's Hertogenbosch, The Netherlands: Katholiek Pedogogisch Centrum.

**SEDL | OERI**

This publication is based on work sponsored wholly, or in part, by the Office of Educational Research & Improvement, U.S. Department of Education, under Contract Number RP91002003. The contents of this publication do not necessarily reflect the views of OERI, the Department, or any other agency of the U. S. Government.

# IX

## *Performance Assessment Examples*

IX-1

# Oregon's Writing Assessment

Oregon Department of Education

The *Oregon Educational Act for the 21st Century*, passed by the 1991 Legislative Assembly, called for a restructured education system to achieve the state's goals to be the "best educated citizens in the nation by the year 2000 and a work force equal to any in the world by the year 2010." This is to be achieved by a massive restructuring, including alternative learning environments, employment of minors, extended school day/year, integration of social services, school choice, site-based decision making, non graded primary classes and a redesign of assessment.

The state recognized that in order to be successful, a systematic approach to restructuring had to occur. It began with the careful definition of key outcomes around which decisions about curriculum, instruction and assessment are being made. The state formally adopted eleven key learning outcomes in 1993. These are six "foundation skills" (thinking, self-directed learning, communication, use of technology, ability to quantify, and ability to collaborate), and five "core applications for living" (ability to deliberate on public issues, understand diversity, interpret human experience, apply science and math, and understand positive health habits).

Students must demonstrate high levels of performance with respect to these key learning outcomes to receive a **Certificate of Initial Mastery** (CIM) by about the end of tenth grade. In order to assess whether students have demonstrated "high levels of performance" the state has adopted "extended definitions" of these outcomes, and is going through a systematic process of developing guidelines for performance tasks, tasks, ways to score performance on the tasks (rubrics) and portfolios. Tasks will include both short "on-demand" performance assessments and more extended projects embedded in classroom work. Students will assemble a portfolio that will contain evidence of growth and attainment of the key outcomes. Content might include on-demand performance assessments and projects as well as other evidence chosen by the student and/or teacher. All portfolio entries must include explanations as to why they were selected. The portfolio must also contain a student self-assessment which includes an explanation of the relationship between portfolio contents and the CIM outcomes and prescribed standards. The portfolio will also contain some performance assessments taken by all students.

Formal assessment of progress toward attainment of the key outcomes will occur in grades 3, 5, 8, and 10. The state is developing performance crite-

---

Reprinted with permission from the Oregon Department of Education, Office of Assessment and Technology, Salem, Oregon.

ria for each key outcome, standards that describe what "adequate progress" looks like at each benchmark year, and sample student work that illustrates adequate progress.

We will illustrate Oregon's approach using its writing assessment, which has been in place since 1985 and provides a model for the other key outcomes.

## Statement of Key Learning Outcome in Writing

Statement of the outcome in writing is part of the *communication* goal: The student will be able to *communicate through reading, writing, speaking, and listening, and through an integrated use of visual forms such as symbols and graphic images.*

This goal is elaborated in a series of expanded definitions. The writing definition says:

> Students will use writing as a tool for learning and self-reflection and to convey meaning through a variety of written forms (e.g., stories, essays, journals, technical reports). An effective writer is able to:
>
> - demonstrate control of ideas and content, organization, voice, word choice, sentence fluency and conventions
> - evaluate and monitor personal writing, using a variety of strategies to produce coherent and mechanically correct final products.

## Task Guidelines

Each task must identify the following:

1. Linkage between the task and rubric(s)
2. Prerequisite content, processes, and strategies which students must know before they begin the task
3. Key curricular content, concepts, issues, and problems that are related to the task and outcome
4. Orientation activity that models common terms, contexts, and background information (to level the playing field)
5. Instructional pieces or mini-lessons that are required for the student to have success during the task
6. Specific foundation outcomes to be assessed and at what point in the task
7. Provisions for assessing individual performance in group work.

Tasks must provide all students an opportunity to learn.

1. The task must be accessible to all students and allow for individual learning styles and needs.
2. The task should allow students to choose from a variety of forms (e.g., oral presentations, dramatic performances, video productions, visual representations, written products) to demonstrate proficiency.
3. The task is a fair assessment of what the student knows and can do while not dependent on unique background knowledge, individual circumstances or characteristics (e.g., locale, cultural heritage, gender, specific disabilities or socioeconomic status).
4. The task provides opportunities for students to explore issues from diverse perspectives (e.g., gender, cultural, regional, socioeconomic) where appropriate.

The task is interesting and engaging to students.

1. The task is developmentally appropriate.
2. The task is written in language that is understandable and clear.
3. The task, where appropriate, gives the student a choice of topic.
4. The task should be approachable and expandable to different curricular areas.

Tasks must have options for personalization.

1. The task requires students to apply interpretation and judgment.
2. The task requires students to engage in self-reflection.

Tasks must have options for transfer to greater life experiences.

1. The task must include reflection of the impact on a variety of life experiences (school, family, community, planet).
2. Students have knowledge of the audience prior to the task.
3. The task is similar to the real-world, authentic tasks that students are likely to face in their life today and in the future.

Questions to ask yourself as you begin to write the task:

1. How does this task fit into my instructional program?
2. How well does the task match the CIM outcome?
3. Is the major theme or curricular purpose of the task obvious to the student?
4. Is this task adding to the variety of content strands assessed?

## Sample Task

The state has been conducting on-demand performance assessments in writing since 1985. Students are randomly assigned to one of five types of writing: persuasive, expository, imaginative, description, or narrative. Students have their choice of one of two prompts. Prompts were written to allow personalization and a range of interpretations. They were written so as not to require any specialized knowledge. An attempt has been made to simulate the writing process—students have three 45-minute sessions in which to write. Students are encouraged to review the criteria for scoring as they write. At the end students answer questions about their attitudes toward writing and the type of writing experiences they have had.

A sample writing booklet for Grade 11 descriptive writing is reproduced in **Sample 1**. The corresponding scan sheet is provided in **Sample 2**.

## Performance Criteria

Each student essay is scored by two raters on six dimensions: ideas, organization, voice, word choice, sentence fluency, and conventions. A scoring guide (rubric) is shown in **Sample 3**. The criteria are designed to define and provide more detail on the expanded definition of what is meant by "communicating well in writing." In actuality, the expanded definition was informed by the performance criteria. The experience obtained in assessing writing was essential in sharpening and clarifying the definition. (There are also scoring guides for the various purposes or modes of writing—persuasive, narrative, descriptive, expository and imaginative. These are used to judge how well student essays fulfill their purpose. The scoring guides are not included in this paper.)

## Reports

Several types of reports are prepared. The first kind, illustrated in **Sample 4** (Oregon Department of Education, 1993), provides samples of student essays at each grade level that illustrate strong, middle and weak performance on each dimension that is scored. **Sample 4** contains strong and weak essays for the descriptive prompt on teachers provided in **Sample 1**.

A second type of report provides average scores for various grade levels, tracks changes over time, and provides recommendations for staff development, instruction, and state action. An example is provided in **Sample 5** (Oregon Department of Education, 1989).

Other reports are: Individual Student (**Sample 6**), classroom (**Sample 7**) and district (**Samples 8** thru **11**).

For more information contact:

Barbara Wolf
Oregon Department of Education
155 Capitol Street N.E.
Salem, OR 97310
(503) 378-8004

### REFERENCES

Oregon Department of Education (1989). *Oregon Statewide Writing Assessment, 1991 and 1992*. Oregon Department of Education, Salem, Oregon 97310.

Oregon Department of Education (1993). *Oregon Statewide Writing Assessment, 1991 and 1992*. Oregon Department of Education, Salem, 97310.

SAMPLE 1

# OREGON Statewide Writing Assessment Booklet

Student Name

Teacher Name

School Name

District Name

1993

Grade 11

## DIRECTIONS

Starting today, you will take part in a *writing assessment*. This means you will write a paper that will be read and scored by two readers other than your teacher. You will have time on three different days--today, tomorrow and the next day--to work on your paper.

### STEP 1: PREWRITING

Look at the next page to find the topics you can write about. Choose one of these topics. When you have chosen, circle your choice. Use a separate sheet of paper to list ideas or do some other prewriting BEFORE you begin to write your rough draft.

### STEP 2: WRITING THE ROUGH DRAFT

Begin writing your rough draft on NOTEBOOK PAPER when you finish prewriting. Be sure to write on the topic you circled. Remember that your *final copy* will need to fit into the Writing Assessment Booklet. This booklet has two pages for your final copy.

### STEP 3: REVISING AND EDITING

When your rough draft is finished, you can spend some time revising and editing. Use the Guide to Revision (on page 2) to help you. It will tell you the main things to look for when you revise and edit.

### STEP 4: RECOPYING AND PROOFREADING

When you are done revising and editing, recopy your paper into the Writing Assessment Booklet. Please use a pen. Your final copy must fit onto the two pages provided. Make your copy as *neat* as you can so that it is easy for others to read. When you finish, make sure your name and your teacher's name are on the front of the booklet. When you finish writing you will also be given a computer "answer sheet" to fill in. Complete all information in sections 1-7 on this sheet. When done, turn in all materials to your teacher.

Thank you for sharing a sample of your writing!

SAMPLE 1—*(continued)*

**Choose only one of these topics to write about. Circle the one you choose:**

1. Almost everyone has had at least one teacher who is hard to forget. Think of a teacher you have had that you will remember for a long time. DESCRIBE that teacher so clearly that your reader will know just what that teacher is like and what makes him or her so hard to forget.

OR

2. Everyone has had some type of work experience, whether it was at home (building or repairing something, babysitting, yard or garden work, etc.) or at a workplace for pay. DESCRIBE a work experience you have had so that a reader can visualize your work environment and understand how it felt to be there.

## GUIDE TO REVISION

**Use the checklist below to help you revise your rough draft before you copy it into this Writing Booklet.**

### IDEAS AND CONTENT

[ ] My paper has a clear purpose or makes a point.
[ ] I choose clear details and examples to help the reader understand my message
[ ] I stick to the main idea. I leave out details that do not matter.
[ ] I am writing about something I know.

### ORGANIZATION

[ ] My introduction would make a reader want to keep reading.
[ ] I tell things in an order that makes sense.
[ ] Details in my paper go together.
[ ] My paper ends in a good spot. It doesn't stop suddenly or drag on too long.

### VOICE

[ ] My writing shows what I really think and feel.
[ ] I like what I've written.
[ ] My writing sounds like me, and not like someone else.
[ ] I have thought about my reader. I have tried to make my writing clear to the reader.

### WORD CHOICE

[ ] I chose words that will help make my meaning clear.
[ ] My words paint a picture in the reader's mind.
[ ] I have tried to find my own way to say things.
[ ] Sometimes I have tried to say something in a new or different way.

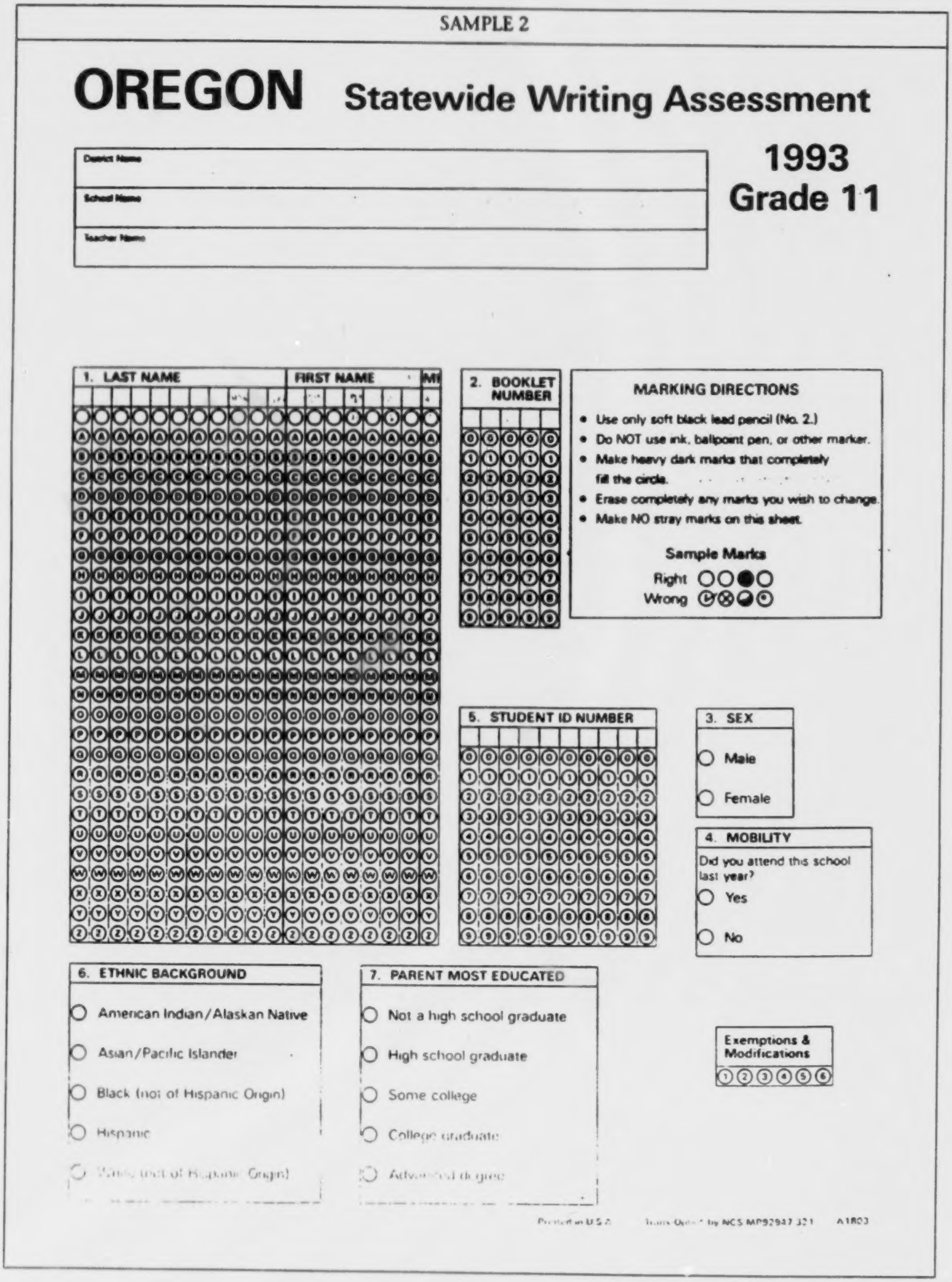
### SENTENCE FLUENCY

[ ] My sentences make sense. They are clear.
[ ] Some sentences are longer than others.
[ ] Sentences begin in different ways. (They do NOT all begin with the same words.)
[ ] My paper would be easy to read out loud.

### CONVENTIONS

[ ] My paragraphs begin in the right spots.
[ ] My punctuation is correct.
[ ] I use capital letters on names of people, places or things.
[ ] I use capital letters to begin sentences.
[ ] My spelling is correct.
[ ] I have proofread my paper.

SAMPLE 2



**OREGON** Statewide Writing Assessment

1993
Grade 11

District Name

School Name

Teacher Name

1. LAST NAME | FIRST NAME | MI

2. BOOKLET NUMBER

MARKING DIRECTIONS

- Use only soft black lead pencil (No. 2.)
- Do NOT use ink, ballpoint pen, or other marker.
- Make heavy dark marks that completely fill the circle.
- Erase completely any marks you wish to change.
- Make NO stray marks on this sheet.

Sample Marks

Right ○○●○
Wrong

5. STUDENT ID NUMBER

3. SEX

○ Male

○ Female

4. MOBILITY

Did you attend this school last year?

○ Yes

○ No

6. ETHNIC BACKGROUND

○ American Indian/Alaskan Native

○ Asian/Pacific Islander

○ Black (not of Hispanic Origin)

○ Hispanic

○ White (not of Hispanic Origin)

7. PARENT MOST EDUCATED

○ Not a high school graduate

○ High school graduate

○ Some college

○ College graduate

○ Advanced degree

Exemptions & Modifications
① ② ③ ④ ⑤ ⑥

Printed in U.S.A. Trans-Optic® by NCS MP92947 321 A1803

BEST COPY AVAILABLE

SAMPLE 2—*(continued)*

# STUDENT WRITER'S QUESTIONS

**Please answer these questions after you finish writing. We'd like to know a little more about how you write and revise. If you have time when you finish writing, please answer these questions for us. This is not a test, and there are no right or wrong answers. Be as honest as you can.**

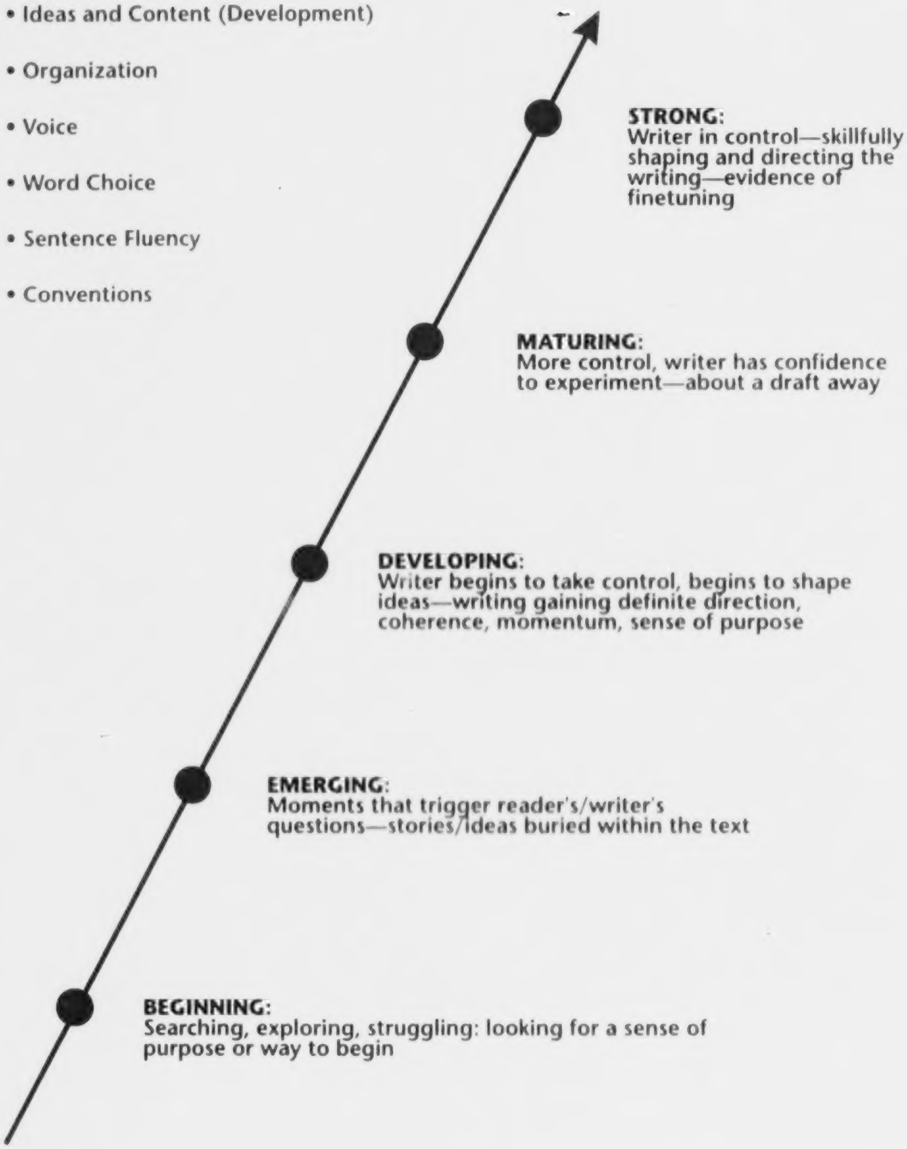
1. **How much did you like writing on this topic?**
   - ○ Very much ○ Not much
   - ○ Somewhat ○ Not at all
2. **How much do you usually like to write?**
   - ○ Very much ○ Not much
   - ○ Somewhat ○ Not at all
3. **What kind of prewriting did you do? Mark the main ones.**
   - ○ Listing words or outlining.
   - ○ Clustering or networking.
   - ○ Thinking about the topic.
   - ○ Talking with a friend or family member outside of class.
   - ○ Something different from the things listed.
   - ○ I do not understand what you mean by prewriting.
4. **Did you write a rough draft?**
   - ○ Yes
   - ○ No
   - ○ I do not understand what a rough draft is.
5. **Which of these things did you do in revising your paper? Mark the main ones.**
   - ○ I read my paper over to myself.
   - ○ I took some things out.
   - ○ I put more information in.
   - ○ I moved things around.
   - ○ I recopied my paper just the way it was without making any changes.
   - ○ I did NO revising.
   - ○ I do not understand what you mean by revising.
6. **Which of these things did you do in editing your paper? Mark the main ones.**
   - ○ I used a dictionary or thesaurus.
   - ○ I read my paper over carefully.
   - ○ I fixed mistakes I noticed in the writing.
   - ○ I tried to make the paper neater.
   - ○ I recopied the paper just the way it was without making any changes.
   - ○ I did NO editing.
   - ○ I do not understand what you mean by editing.

| Mode: Ⓓ Ⓔ Ⓘ Ⓝ Ⓟ | Prompt ① ② | For Official Use Only | | | | | |
|---|---|---|---|---|---|---|---|
| I | O | V | W | S | C | [illegible] | SC |
| ①②③④⑤ | ①②③④⑤ | | | | | | |
| ①②③④⑤ | ①②③④⑤ | | | | | | |
| ①②③④⑤ | ①②③④⑤ | | | | | | |

BEST COPY AVAILABLE

## SAMPLE 3
## Analytical Scoring Guide: Traits

- Ideas and Content (Development)
- Organization
- Voice
- Word Choice
- Sentence Fluency
- Conventions



**STRONG:**
Writer in control—skillfully shaping and directing the writing—evidence of finetuning

**MATURING:**
More control, writer has confidence to experiment—about a draft away

**DEVELOPING:**
Writer begins to take control, begins to shape ideas—writing gaining definite direction, coherence, momentum, sense of purpose

**EMERGING:**
Moments that trigger reader's/writer's questions—stories/ideas buried within the text

**BEGINNING:**
Searching, exploring, struggling: looking for a sense of purpose or way to begin

Developed by the Northwest Regional Educational Laboratory, January 1992. This scoring guide is an updated version of the one that appears in Spandel and Stiggins, *Creating Writers*. Addison-Wesley: 1990. The original guide was developed by teachers from Beaverton, Oregon School District in 1984. The Laboratory gratefully acknowledges the contributions of more than 5,000 teachers whose shared insights and comments have gone into this revision.

SAMPLE 3

## Analytical Scoring Guide: Traits

### IDEAS AND CONTENT (Development)

**5: Strong. This paper is clear, focused, and interesting. It holds the reader's attention. Relevant anecdotes and details enrich the central theme or story line. Ideas are fresh and original.**

- The writer seems to be writing from experience and shows insight: a good sense of how events unfold, how people respond to life and to each other.
- Supporting, relevant, telling details give the reader important information that he or she could not personally bring to the text.
- The writing has balance: Main ideas stand out.
- The writer seems in control and develops the topic in an enlightening, entertaining way.
- The writer works with and shapes ideas, making connections and sharing insights.

**3: Developing. The paper is clear and focused. The topic shows promise, even though development is still limited, sketchy or general.**

- The writer is beginning to define the topic, but is not there yet. It is pretty easy to see where the writer is headed, though more formation is needed to "fill in the blanks."
- The writer does seem to be writing from experience, but has some trouble going from general observations to specifics.
- Ideas are reasonably clear and purposeful, even though they may not be explicit, detailed, personalized, or expanded to show in-depth understanding.
- Support is attempted, but doesn't go far enough yet in expanding, clarifying, or adding new insights.

**1: Beginning. As yet, the paper has no clear sense of purpose or central theme. To extract meaning from the text, the reader must make inferences based on sketchy details. More than one of the following problems is likely to be evident:**

- Information is very limited or unclear.
- The text is very repetitious, or reads like a collection of random thoughts from which no central theme emerges.
- Everything seems as important as everything else; the reader has a hard time sifting out what's critical.
- The writer has not yet begun to define the topic in a meaningful or personal way.
- The writer may still be in search of a real topic, or sense of direction to guide development.

### ORGANIZATION

**5: Strong. The organization enhances and showcases the central idea or theme. The order, structure, or presentation is compelling and moves the reader through the text.**

- Details seem to fit where they're placed; sequencing is logical and effective.
- An inviting introduction draws the reader in and a satisfying conclusion leaves the reader with a sense of resolution.
- Pacing is very well controlled; the writer delivers needed information at just the right moment, then moves on.
- Transitions are smooth and weave the separate threads of meaning into one cohesive whole.
- Organization flows so smoothly the reader hardly thinks about it.

**3: Developing. The organizational structure is strong enough to move the reader from point to point without undue confusion.**

- The paper has a recognizable introduction and conclusion. The introduction may not create a strong sense of anticipation; the conclusion may not leave the reader with a satisfying sense of resolution.
- Sequencing is usually logical. It may sometimes be too obvious, or otherwise ineffective.
- Pacing is fairly well controlled, though the writer sometimes spurts ahead too quickly or spends too much time on the obvious.
- Transitions often work well; at times though, connections between ideas are fuzzy or call for inferences.
- Despite a few problems, the organization does not seriously get in the way of the main point or storyline.

SAMPLE 3—*(continued)*

**1: Beginning. The writing lacks a clear sense of direction. Ideas, details or events seem strung together in a random, haphazard fashion—or else there is no identifiable internal structure at all. More than one of the following problems is likely to be evident:**

- The writer has not yet drafted a real lead or conclusion.
- Transitions are not yet clearly defined; connections between ideas seem confusing or incomplete.
- Sequencing, if it exists, needs work.
- Pacing feels awkward, with lots of time spent on minor details or big, hard-to-follow leaps from point to point.
- Lack of organization makes it hard for the reader to get a grip on the main point or story line.

### VOICE

**5: Strong. The writer speaks directly to the reader in a way that is individualistic, expressive, and engaging. Clearly, the writer is involved in the text and is writing to be read.**

- The paper is honest and written from the heart. It has the ring of conviction.
- The language is natural yet provocative; it brings the topic to life.
- The reader feels a strong sense of interaction with the writer and senses the person behind the words.
- The projected tone and voice give flavor to the writer's message and seem very appropriate for the purpose and audience.

**3: Developing. The writer seems sincere, but not genuinely engaged, committed, or involved. The result is pleasant and sometimes even personable, but short of compelling.**

- The writing communicates in an earnest, pleasing manner. Moments here and there amuse, surprise, delight or move the reader.
- Voice may emerge strongly on occasion, then retreat behind general, vague, tentative, or abstract language.
- The writing hides as much of the writer as it reveals.
- The writer seems aware of an audience, but often to weight words carefully, to stands at a distance, and to avoid risk.

**1: Beginning. The writer seems indifferent, uninvolved or distanced from the topic and/or the audience. As a result, the writing is flat, lifeless or mechanical; depending on the topic, it may be overly technical or jargonistic. More than one of the following problems is likely to be evident:**

- The reader has a hard time sensing the writer behind the words. The writer does not seem to reach out to an audience, or make use of voice to connect with that audience.
- The writer speaks in a kind of monotone that tends to flatten all potential highs and lows of the message.
- The writing communicates on a functional level, with no apparent attempt to move or involve the reader.
- The writer is not yet sufficiently engaged or at home with the topic to take risks or share him-/herself

### WORD CHOICE

**5: Strong. Words convey the intended message in an interesting, precise, and natural way. The writing is full and rich, yet concise.**

- Words are specific and accurate; they seem just right.
- Imagery is strong.
- Powerful verbs give the writing energy.
- Striking words and phrases often catch the reader's eye, but the language is natural and never overdone.
- Expression is fresh and appealing; slang is used sparingly.

**3: Developing. The language is functional, even if it lacks punch; it does get the message across.**

- Words are almost always correct and adequate (though not necessarily precise); it is easy to understand what the writer means.
- Familiar words and phrases communicate, but rarely capture the reader's imagination. The writer seems reluctant to stretch.
- The writer usually avoids experimenting; however, the paper may have one or two fine moments.
- Attempts at colorful language often come close to the mark, but may seem overdone or out of place.
- A few energetic verbs liven things up now and then; the reader yearns for more.
- The writer may lean a little on redundancy, or slip in a cliché but never relies on these crutches to the point of annoyance.

## SAMPLE 3—*(continued)*

**1: Beginning. The writer struggles with a limited vocabulary, searching for words to convey meaning. More than one of the following problems is likely to be evident:**

- Language is so vague and abstract (e.g., it was a fun time, it was nice and stuff) that only the most general message comes through.
- Persistent redundancy clouds the message and distracts the reader.
- Clichés or jargon serves as a crutch.
- Words are used incorrectly in more than one or two cases, sometimes making the message hard to decipher.
- The writer is not yet selecting words that would help the reader to a better understanding.

### SENTENCE FLUENCY

**5: Strong. The writing has an easy flow and rhythm when read aloud. Sentences are well built, with consistently strong and varied structure that makes expressive oral reading easy and enjoyable.**

- Sentence structure reflects logic and sense, helping to show how ideas relate. Purposeful sentence beginnings guide the reader readily from one sentence to another.
- The writing sounds natural and fluent; it glides along with one sentence flowing effortlessly into the next.
- Sentences display an effective combination of power and grace.
- Variation in sentence structure and length adds interest to the text.
- Fragments, if used at all, work well.
- Dialogue, if used, sounds natural.

**3: Developing. The text hums along efficiently for the most part, though it may lack a certain rhythm or grace. It tends to be more pleasant or businesslike than musical, more mechanical than fluid.**

- The writer shows good control over simple sentence structure, more variable control over complex sentence structure.
- Sentences may not seem skillfully crafted or musical, but they are grammatical and solid. They hang together. They get the job done.
- The writer may tend to favor a particular pattern (e.g., subject-verb, subject-verb), but there is at least *some* variation in sentence length and structure (Sentence beginnings are not all alike).
- The reader sometimes has to hunt for clues (e.g., connecting words like *however, therefore, naturally, on the other hand, to be specific, for example, next, first of all, later, still;* etc.) that show how one sentence leads into the next.
- Some parts of the text invite expressive oral reading; others may be a little stiff, choppy or awkward. Overall, though, it's pretty easy to read this paper aloud if you practice.

**1: Beginning. The paper is difficult to follow or read aloud. Most sentences tend to be choppy, incomplete, rambling, or awkward; they need work. More than one of the following problems is likely to be evident:**

- Sentences do not sound natural, the way someone might speak. Word patterns are often jarring or irregular, forcing the reader to pause or read over.
- Sentence structure tends to obscure meaning, rather than showing the reader how ideas relate.
- Word patterns are very monotonous (e.g., subject-verb, subject-verb-object). There is little or no real variety in length or structure.
- Sentences may be very choppy. Or, words may run together in one giant "sentence" linked by "*and's*" or other connectives.
- The text does not invite expressive oral reading.

### CONVENTIONS

**5: Strong. The writer demonstrates a good grasp of standard writing conventions (e.g., grammar, capitalization, punctuation, usage, spelling, paragraphing) and uses them effectively to enhance readability. Errors tend to be so few and minor the reader can easily skim right over them unless specifically searching for them.**

- Paragraphing tends to be sound and to reinforce the organizational structure.
- Grammar and usage are correct and contribute to clarity and style.
- Punctuation is smooth and guides the reader through the text.
- Spelling is generally correct, even on more difficult words.
- The writer may manipulate conventions—particularly grammar—for stylistic effect.
- The writing is sufficiently long and complex to allow the writer to show skill in using a wide range of conventions.
- Only light editing would be required to polish the text for publication.

SAMPLE 3—*(continued)*

**3. Developing. The writer shows reasonable control over a limited range of standard writing conventions. However, the paper would require moderate editing prior to publication. Errors are numerous or serious enough to be somewhat distracting, but the writer also handles some conventions well.**

- Spelling is usually correct (or reasonably phonetic) on common words.
- Terminal (end-of-sentence) punctuation is almost always correct; internal punctuation (commas, apostrophes, semicolons) may be incorrect or missing.
- Problems with grammar or usage are not serious enough to distort meaning.
- Paragraphing is attempted. Paragraphs sometimes run together or begin in the wrong places.
- The paper seems to reflect light, but not extensive or thorough, editing.

**1: Beginning. Errors in spelling, punctuation, usage and grammar, capitalization and/or paragraphing repeatedly distract the reader and make the text difficult to read. More than one of the following problems is likely to be evident:**

- The reader must read once to decode, then again for meaning.
- Spelling errors are frequent, even on common words.
- Punctuation (including terminal punctuation) is often missing or incorrect.
- Paragraphing is missing, irregular, or so frequent (e.g., every sentence) that it does not relate to organization of the text.
- Errors in grammar and usage are very noticeable, and may affect meaning.
- Extensive editing would be required to polish the text for publication.

SAMPLE 4

## 11th Grade Student Papers—Modes

**Descriptive—High.** The teacher in this essay gives her students Jolly Rancher candy and continues teaching right through the fire drills. Specific, well-chosen details bring this memorable character to life for us.

. . . . . . .

Miss Rodriquez, a short, bubbly woman with red hair, bounced into Spanish class on the first day of school. She laughed and smiled, and spoke to the class in Spanish, although no one really understood what she was saying. However, she assured the students that by the end of the year, they would be able to understand. Then she opened the door to a closet in her room and gave everyone a Jolly Rancher.

Miss Rodriquez' goal for every student in her classes was to have each individual learn as much Spanish as possible. She wanted her students to know not only the grammar, but also the culture, and the idiosyncrasies of the Spanish people. However, she wanted people to have fun while learning. She seized every opportunity to make every lesson new and exciting. Each day after class, the students walked out wanting to learn more Spanish. She wasn't afraid to do anything. Many of her lessons seemed silly and ridiculous, but they helped people to learn better. She would sing or dance in front of the class, and create chants to emphasize a point. Not until after class did the realization come that something had been learned from her crazy endeavors. One time the fire alarm sounded, and, as usual, the student body had to gather outside. She was in the middle of teaching a difficult lesson on a grammatical structure, and she didn't want the class to forget what she had been talking about. When the whole class was out of the building, she gathered everyone together and continued the lesson. The rest of the school gave the class some rather strange looks, but she kept right on teaching.

In addition to her academic side, Miss Rodriquez cared about everyone as a person. She wanted people to learn about life, how to succeed, accomplish goals, and realize dreams. If someone ever needed help with school or with life in general, she was there to listen. Her room was always open during lunch and after school. People could go down to her room to talk or to have a great time. She boosted everyone's self-confidence, and made people realize that they could achieve what they set out to achieve. To all her first-year classes, she showed the movie of Don Quixote de La Mancha. The movie was a story of an old man who put his dream of making the world better into reality. No matter what the rest of the world said, he pursued his dream. After the movie was over, she lovingly told the class that they could, and should, also have dreams. She reminded the students that they could be like Don Quixote and not to let anyone spoil their dreams. To a young seventh-grader just entering the start of the "real" world, those were encouraging and motivating words. Also, she challenged her students to achieve greater heights. Many times she told them that they were doing the same work as high school and college students, and that they could do better. She encouraged all of her students to learn verb tenses and vocabulary beyond what the course designated.

Miss Rodriquez was dedicated not only to her Spanish students, but to the rest of the school as well. She was in charge of the Spanish Club, and she put in a lot of time and effort to organize meetings and parties. Although she was the director, she let the members decide on almost everything. She extended the membership to include students not enrolled in a Spanish class. Spanish club gatherings were fun times when friends could talk and meet new people. She enjoyed providing a place for teenagers to relax and be removed from the pressures of life.

Miss Rodriquez was a memorable teacher. She put a tremendous amount of love and care into her teaching. While in her class, people were considered to be human beings, not just another student shuffling through yet another class. Whenever she was needed, she was always there ready to listen and help. Learning, in her class, was not only interesting but also exciting. She was considerate, concerned, and lively. Many memories were related and lessons learned in her classes that will not be forgotten by her students.

SAMPLE 4—*(continued)*

**Descriptive—Low.** It is hard to get more than the most general sense of who this teacher is and what she is like. This writer has not yet given us the details needed to create either strong mental images or a real feeling for how it was to be there.

.......

That is correct every body does have a teacher that they will never forget.

The teacher, that I am going to describe is a very special teacher this teacher has been working with me every since my freshman year, and every since then I have admired her to the T. Della Waller has inspired me since my freshman year. When any body first meets someone they always get there stereo type well I had mine. This teacher was all most to clear on class information. I could tell right off the bat that this was going to be a wonderful four years of high school. Della had a heart of gold, A mind of a genuis. My stereo type was like a bohemeing type of hippy of cource thats good other wise I wouln'nt be writting about her. A student that comes into this class will not, I say will not leave here if della thinks he or she is not clear on her teachings. Della and I have had are little falling outs, every body has there difernces and no body is perfect. Della picked her way to succses and it is more than a job it is her life, and it's my job to play that special roll in her life. . . . A student.

The teacher I can't forget about is a teacher named Mr. Stubbs. Mr. Stubbs is a teacher that has not grownup. He was a welding and meltles shop teacher? He would like to miss around in the shop. Like take very hot pices of meltle and sticking them in your coverals.

What we would do to get back is to flip up his wilding sheld when he wes welding. Sometimes we would ask him to look at a weld we had just don while it was still hot and hand it to him. But he had a scrious side to him. When there was work to be done he made us get it done on time. There was a teacher that didn't like Mr. Stubbs. That teacher would tell him that he wes mot dowing his job and he would come in and see if he was doing his job. When it wus mot his job to do that.

But Mr. Stubbs was a good teacher. He tought us what we meeded to mow about.

SAMPLE 5

## Chapter 1:

### The Oregon Writing Assessment Interpretive Panel— Comments and Recommendations

The Oregon Writing Assessment Interpretive Panel met at the Oregon Department of Education in Salem on May 5, 1989, to review the results of the 1989 statewide writing assessment. They met for three purposes:

- To help lay and professional audiences understand the results more fully.
- To identify key issues for consideration in the future.
- To offer recommendations concerning actions that writing teachers and that the Oregon Department of Education might take to improve the instruction of writing.

During their meeting, the panel members reviewed the prompts, the methods by which scores were assigned and summaries of students' performance relative to all traits. They considered both average performance on all traits, and the score distributions, or percentage of students who had received each possible score (1.0, 1.5, 2.0, etc.) on the five-point scoring scale. (See Appendix D for a list of Panel Members.)

**Findings.** Following are the highlights of the panel's findings.

1. The average performance of eighth grade students from 1987 to 1989 is virtually unchanged across all six traits as Table 1 illustrates:

**Table 1: Rating Means**

| Scale | 1987 | 1989 (n=2868) |
|---|---|---|
| Ideas | 3.160 | 3.084 |
| Organization | 3.009 | 2.937 |
| Voice | 3.267 | 3.204 |
| Word Choice | 3.147 | 3.128 |
| Sentence Fluency | 3.181 | 3.191 |
| Conventions | 3.228 | 3.182 |

2. In three areas of performance—Voice, Sentence Fluency and Conventions—nearly half of the students scored 3.5 or better, indicating more strengths than weaknesses. In fact, in four of the six traits, the number of students for whom strengths outweighed weaknesses exceeded the number for whom weaknesses outweighed strengths. This is an encouraging sign, and underscores the general trend toward improved writing performance since the first analytical assessment in 1985. Though it is too early to draw final conclusions, we may be seeing the results of changing classroom practices: e.g., more time devoted to writing, and more emphasis on the teaching writing as a process.

3. The increased length of student papers this year, together with the relatively high performance in Sentence Fluency, suggests greater fluency in student writing which may indicate that some students, at least, are growing more comfortable with writing.

4. The pattern of performance for Word Choice, as in 1985 and 1987, continues to show a clustering of scores at the 3.0 level, with relatively fewer students showing either great strength or pronounced weakness on this trait. At the same time, however, more students in 1989 showed strengths on Word Choice than showed weaknesses, again suggesting the beginnings of a promising trend. The 3.0 score in Word Choice, which balances strengths and weaknesses, reflects the kind of language that communicates in a routine, functional way. Most students are using language that is reasonably clear and understandable. Not enough students, however, are using language to think, to clarify ideas, to enliven or enrich their description, or to explore new ways of looking at things. The answer is not more vocabulary drills, however. A more effective way to resolve this problem is to surround our students with a language-rich environment, to provide opportunities for them to hear and to read literature, and to respond to that literature in writing and through discussion.

5. The relatively strong performance in Conventions provides encouraging evidence that students are using punctuation, grammar, paragraphing, spelling and capitalization effectively in their writing.

SAMPLE 5—*(continued)*

6. Despite some apparent progress since 1985, scores on the trait of Organization continue to be relatively low in comparison to scores on other traits. Considering the apparent strong connection between this trait and thinking skills, we should provide more opportunities for students to develop strengths in this trait by writing in a wide range of genres, and for varied purposes and audiences. This, in turn, means providing opportunities for students to write in many areas of the curriculum, not only in writing or literature classes. On the encouraging side, we see many students taking a creative approach to organization. Few show dependence on an artificial structure, such as the five-paragraph essay. Instead, the organization in increasing numbers of papers shows a clear and deliberate connection to a strong theme, with a well-thought-out, inviting introduction; smooth and strong transitions; an identifiable climax or "high point"; and a conclusion with a clear, satisfying sense of resolution.

7. Papers which received lower scores on the traits of Ideas and Content and/or Organization often suffered from an over-abundance of trivial detail. Many recounted trips to Disneyland or other vacation adventures in which the writer strove to tell every event of the journey, whether it enhanced the story development or not. The most successful papers tended to focus on specific incidents (e.g., catching a ball at a baseball game, eating a worm alive, flying in a plane for the first time) which stood out vividly in the writer's mind. The writer selected just those details which advanced the storyline or contributed to development of a central idea, and wove the individual element of the larger whole together in a way that brought each incident alive for the reader. Successful papers avoided lists, and shunned generalities. They were specific, rich with detail, and bright with the energy of a writer who took a general topic and made it his or her own.

8. Assessing traits individually has value because it allows us to see where relative strengths and weaknesses in students' writing lie. Instruction on the traits, however, should be given within the context of overall writing skills; in other words, individual traits (e.g., Ideas or Organization) should not be taught in isolation.

9. The large numbers of prewriting sheets and rough drafts included with this year's assessment samples indicated to readers that many student writers are making real use of the writing process, spending time thinking through and rehearsing ideas before writing, then taking time to revise after the first draft. A nonsystematic analysis of the differences between rough and final drafts indicates much evidence of real effort on the part of student writers.

10. Writing assessment is critical to guiding sound decisions about curriculum. This is so because assessment of writing taps more than writing per se; it goes below the surface to measure thinking skills, and thus provides invaluable information about students' ability to make use of what they learn both in and out of school.

**Recommendations: District, School and Teacher Levels**

1. Ensure that teacher inservice reflects an understanding of the instructional methods research says are most productive: e.g., allowing students frequent opportunities to write, giving student writers time to write and reflect and rewrite, allowing students to develop their own topics and helping them learn to do this effectively, encouraging teachers to write with students and to share their writing, encouraging collaborative learning with student writers working in groups (not only to revise writing, but to rehearse and prepare and drafts), focusing on positive responses, distinguishing between revising and editing so that revising is taught as "seeing again," making student writers—not teachers—responsible for editing, teaching editing skills in the context of extended writing and not through isolated drills, integrating writing and other language arts instruction and involving students themselves in the evaluation process.

2. Provide inservice to help teachers create the kind of school and classroom environment that fosters good written expression.

3. Ensure that the traits are known to both parents and students by introducing students to the scoring guide trait by trait, helping students develop a working vocabulary for talking about writing by using the analytical scoring guide in the classroom, using the scoring guide as a basis for parent-teacher conferences, and making classroom criteria for good writing clearly visible (e.g., with posters).

SAMPLE 5—*(continued)*

4. Ensure that the target goals for good writing are clear in every classroom. Each teacher's expectations for writing performance should be shared openly and in writing with all students, and should be explained carefully so that they are well understood.

5. Provide models that show teachers how to work together at the building level to foster better classroom instruction and assessment. Find ways to build a stronger bridge between assessment and instruction by ensuring that any classroom assessment system is based directly on the objectives and goals that drive instruction in that classroom.

6. Encourage students to take more risks in their writing. Reward performance that reflects an individual approach or a new slant on a topic. Encourage students to share their writing with other students often so that they have an opportunity to see that there is more than one way to approach a topic, and that there is value in these differences for they enrich and expand the way we think about a subject.

7. Distinguish between revising (seeing again, reconceptualizing, reshaping or reworking the whole) and editing (proofreading and correcting). Provide teachers with more strategies for teaching both revising and editing effectively, and encourage teachers to make students responsible for revising and editing their own work.

8. Give students many opportunities to read different kinds of good writing—e.g., literature, persuasive writing, technical writing, expository prose, journalistic writing—and to respond to that writing both orally and in their own written words. In addition, give students experience every year in writing several different types of papers: e.g., narrative, expository, descriptive, persuasive, instructional, technical, analytical, journalistic. Students should not write only essays or only stories; the challenge of variety will help students develop both fluency and critical thinking skills.

9. Encourage teachers to share many samples of good writing—both students' writing and other writing—by reading the samples aloud to students and inviting their response.

10. Encourage districts to use analytical writing assessment as a form of teacher inservice, focusing on the six traits in the Department of Education's analytical scoring guide, and training teachers as raters. Teachers who participate in this training not only gain ideas for teaching writing effectively, but also learn ways of making ongoing classroom assessment of writing more effective and more fair to students.

11. The analytical scoring guide has potential as a tool for teaching content teachers who deal with the traits most closely related to thinking and reasoning skills, notably Ideas and Content, Organization and Voice (in the sense of audience awareness).

12. Beware of setting expectations too low. Students can write. The potential has been demonstrated repeatedly, and we must provide increased opportunities for students to build upon the skills they already have.

13. One of the primary values in the scoring guide is that it encourages all teachers who use it to look at writing in the same way—to be consistent about what they value. We can build on this strength by encouraging teachers to work together at the building level, scoring papers, forming support groups, and discussing what they value in writing. In addition, we should ensure that every teacher and student has access to copy of the scoring guide following the assessment to assist in the appropriate interpretation of the scores.

14. Provide more opportunities for student writers to define their own topics and to develop topics in their own way.

**Recommendations: State Level**

1. Increase support for the Oregon Statewide Writing Assessment. Expand the scope of the assessment to include more students at more grade levels, and provide more training opportunities for teachers. If possible, regionalize the training and scoring activities to provide opportunities for teachers throughout the state to participate as readers and to gain the inservice benefits that are part of that experience.

2. Provide increased support to projects such as the Oregon Writing Project and the statewide Oregon Writing Festival which demonstrate a strong commitment to improving writing skills among students and teachers alike.

SAMPLE 5—*(continued)*

3. Follow up assessment results with a statewide teacher survey to determine how many teachers are familiar with and are using components of the writing process, how many are familiar with or using the six analytical traits as a component of instruction or assessment, and how many are using writing as a way of building thinking skills.

4. Recommend to the Teacher Standards and Practices Commission that it require greater emphasis on preparation for the teaching of writing skills within college and university teacher training programs for teachers in all disciplines, and that such preparation include instruction in the fundamentals of sound classroom writing assessment.

5. Encourage ESDs to coordinate activities that support writing instruction and assessment.

6. In future state writing assessments, continue and expand opportunities for teachers from participating districts to serve as raters on the analytical scoring team.

7. Ensure that the State Board of Education, legislators and budget committees have an opportunity to hear from classroom teachers about the fundamental impact that their participation in the statewide writing assessment has had on their knowledge of classroom assessment, their confidence in assessing student writing, their ability to teach writing effectively, and their way of thinking about student writing.

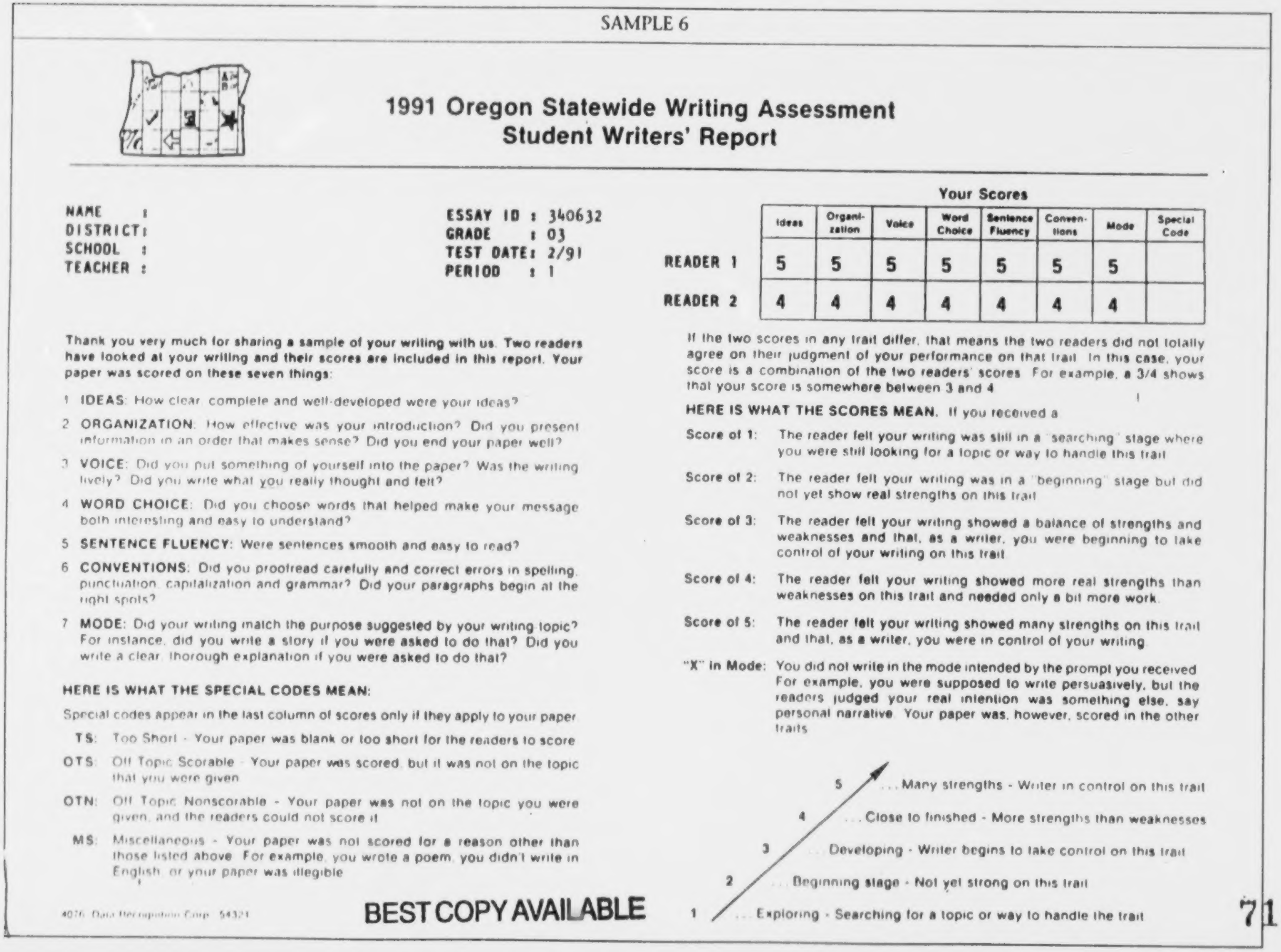
8. Explore the need for specific workshops on these and related topics:

- Effective classroom assessment practices
- Analytical writing assessment
- The link between assessment and instruction
- Teaching to the traits
- Effective conferencing
- Managing peer review
- Using writing to build critical thinking skills
- Assessing writing in the content areas
- Teaching students to revise and edit effectively
- Using word processing in writing instruction
- Effective use of the writing process

In addition, regionalize inservice activities to reach more teachers throughout the state.

9. Identify teachers who are teaching writing effectively, and sponsor regional conferences to offer these teachers opportunities to share their ideas with other teachers of writing.

10. Encourage presentations on the Oregon Statewide Writing Assessment at the Elementary Principals' Conference, COSA, the Secondary Principals' Conference and other forums. Ensure that those who are considering building on district level assessment have an opportunity to hear and understand the benefits of the teachers-teaching-teachers approach to assessment which is the heart and strength of the Oregon analytical model.



SAMPLE 6

## 1991 Oregon Statewide Writing Assessment
## Student Writers' Report

| | |
|---|---|
| **NAME :** | **ESSAY ID :** 340632 |
| **DISTRICT:** | **GRADE :** 03 |
| **SCHOOL :** | **TEST DATE:** 2/91 |
| **TEACHER :** | **PERIOD :** 1 |

**Your Scores**

| | Ideas | Organization | Voice | Word Choice | Sentence Fluency | Conventions | Mode | Special Code |
|---|---|---|---|---|---|---|---|---|
| **READER 1** | 5 | 5 | 5 | 5 | 5 | 5 | 5 | |
| **READER 2** | 4 | 4 | 4 | 4 | 4 | 4 | 4 | |

**Thank you very much for sharing a sample of your writing with us. Two readers have looked at your writing and their scores are included in this report. Your paper was scored on these seven things:**

1 **IDEAS**: How clear, complete and well-developed were your ideas?

2 **ORGANIZATION**: How effective was your introduction? Did you present information in an order that makes sense? Did you end your paper well?

3 **VOICE**: Did you put something of yourself into the paper? Was the writing lively? Did you write what you really thought and felt?

4 **WORD CHOICE**: Did you choose words that helped make your message both interesting and easy to understand?

5 **SENTENCE FLUENCY**: Were sentences smooth and easy to read?

6 **CONVENTIONS**: Did you proofread carefully and correct errors in spelling, punctuation, capitalization and grammar? Did your paragraphs begin at the right spots?

7 **MODE**: Did your writing match the purpose suggested by your writing topic? For instance, did you write a story if you were asked to do that? Did you write a clear, thorough explanation if you were asked to do that?

**HERE IS WHAT THE SPECIAL CODES MEAN:**

Special codes appear in the last column of scores only if they apply to your paper

**TS**: Too Short - Your paper was blank or too short for the readers to score

**OTS**: Off Topic Scorable - Your paper was scored, but it was not on the topic that you were given

**OTN**: Off Topic Nonscorable - Your paper was not on the topic you were given, and the readers could not score it

**MS**: Miscellaneous - Your paper was not scored for a reason other than those listed above. For example, you wrote a poem, you didn't write in English, or your paper was illegible

If the two scores in any trait differ, that means the two readers did not totally agree on their judgment of your performance on that trait. In this case, your score is a combination of the two readers' scores. For example, a 3/4 shows that your score is somewhere between 3 and 4.

**HERE IS WHAT THE SCORES MEAN.** If you received a

**Score of 1:** The reader felt your writing was still in a "searching" stage where you were still looking for a topic or way to handle this trait.

**Score of 2:** The reader felt your writing was in a "beginning" stage but did not yet show real strengths on this trait.

**Score of 3:** The reader felt your writing showed a balance of strengths and weaknesses and that, as a writer, you were beginning to take control of your writing on this trait.

**Score of 4:** The reader felt your writing showed more real strengths than weaknesses on this trait and needed only a bit more work.

**Score of 5:** The reader felt your writing showed many strengths on this trait and that, as a writer, you were in control of your writing.

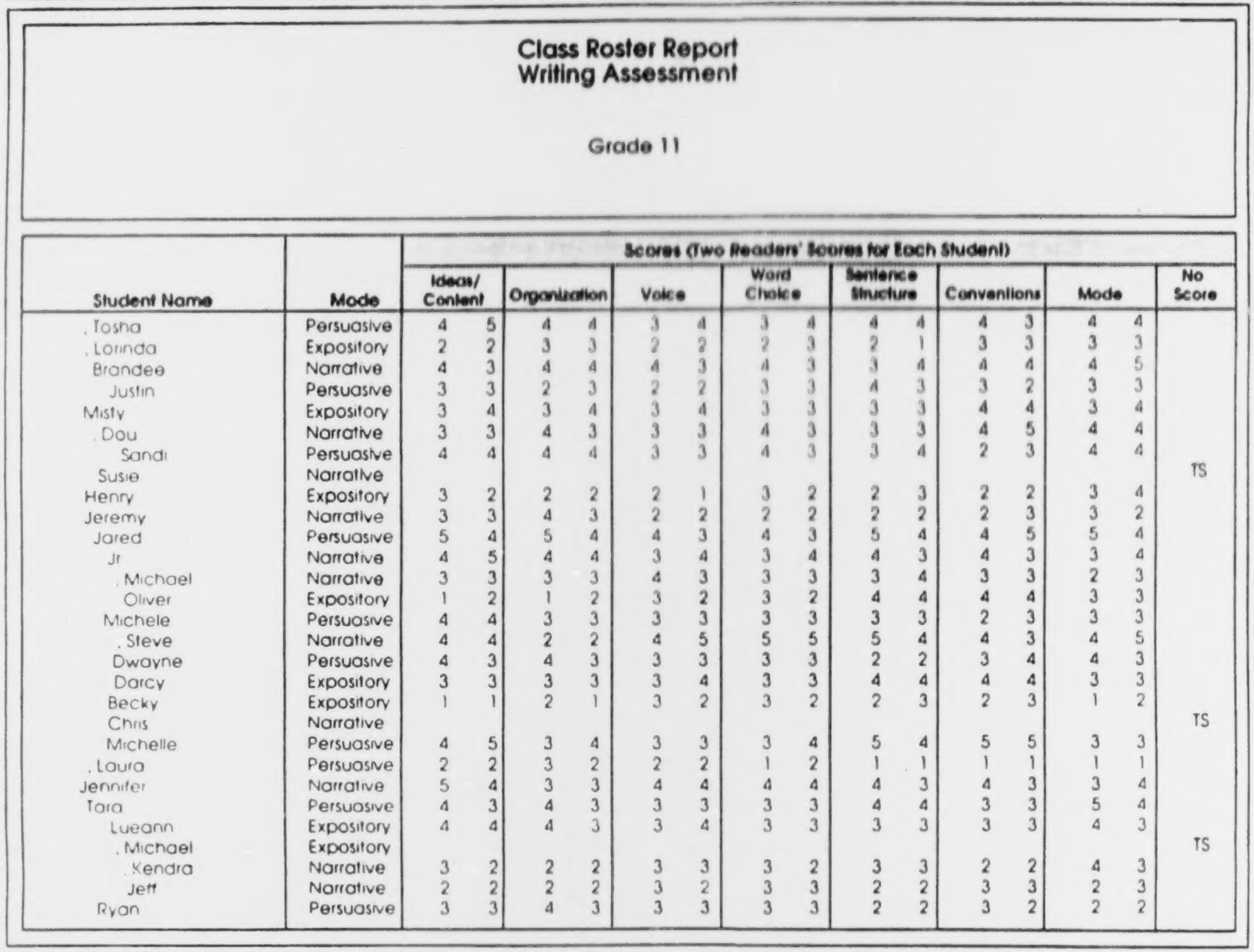
**"X" in Mode:** You did not write in the mode intended by the prompt you received. For example, you were supposed to write persuasively, but the readers judged your real intention was something else, say personal narrative. Your paper was, however, scored in the other traits.

5 ... Many strengths - Writer in control on this trait
4 ... Close to finished - More strengths than weaknesses
3 ... Developing - Writer begins to take control on this trait
2 ... Beginning stage - Not yet strong on this trait
1 ... Exploring - Searching for a topic or way to handle the trait

4076 Data Recognition Corp 54321

BEST COPY AVAILABLE

## SAMPLE 7



### Class Roster Report
### Writing Assessment

Grade 11

| Student Name | Mode | Scores (Two Readers' Scores for Each Student) | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Ideas/ Content | | Organization | | Voice | | Word Choice | | Sentence Structure | | Conventions | | Mode | | No Score |
| , Tosha | Persuasive | 4 | 5 | 4 | 4 | 3 | 4 | 3 | 4 | 4 | 4 | 4 | 3 | 4 | 4 | |
| , Lorinda | Expository | 2 | 2 | 3 | 3 | 2 | 2 | 2 | 3 | 2 | 1 | 3 | 3 | 3 | 3 | |
| Brandee | Narrative | 4 | 3 | 4 | 4 | 4 | 3 | 4 | 3 | 3 | 4 | 4 | 4 | 4 | 5 | |
| Justin | Persuasive | 3 | 3 | 2 | 3 | 2 | 2 | 3 | 3 | 4 | 3 | 3 | 2 | 3 | 3 | |
| Misty | Expository | 3 | 4 | 3 | 4 | 3 | 4 | 3 | 3 | 3 | 3 | 4 | 4 | 3 | 4 | |
| . Dou | Narrative | 3 | 3 | 4 | 3 | 3 | 3 | 4 | 3 | 3 | 3 | 4 | 5 | 4 | 4 | |
| Sandi | Persuasive | 4 | 4 | 4 | 4 | 3 | 3 | 4 | 3 | 3 | 4 | 2 | 3 | 4 | 4 | |
| Susie | Narrative | | | | | | | | | | | | | | | TS |
| Henry | Expository | 3 | 2 | 2 | 2 | 2 | 1 | 3 | 2 | 2 | 3 | 2 | 2 | 3 | 4 | |
| Jeremy | Narrative | 3 | 3 | 4 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 2 | |
| Jared | Persuasive | 5 | 4 | 5 | 4 | 4 | 3 | 4 | 3 | 5 | 4 | 4 | 5 | 5 | 4 | |
| Jr | Narrative | 4 | 5 | 4 | 4 | 3 | 4 | 3 | 4 | 4 | 3 | 4 | 3 | 3 | 4 | |
| . Michael | Narrative | 3 | 3 | 3 | 3 | 4 | 3 | 3 | 3 | 3 | 4 | 3 | 3 | 2 | 3 | |
| Oliver | Expository | 1 | 2 | 1 | 2 | 3 | 2 | 3 | 2 | 4 | 4 | 4 | 4 | 3 | 3 | |
| Michele | Persuasive | 4 | 4 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 3 | 3 | 3 | |
| . Steve | Narrative | 4 | 4 | 2 | 2 | 4 | 5 | 5 | 5 | 5 | 4 | 4 | 3 | 4 | 5 | |
| Dwayne | Persuasive | 4 | 3 | 4 | 3 | 3 | 3 | 3 | 3 | 2 | 2 | 3 | 4 | 4 | 3 | |
| Darcy | Expository | 3 | 3 | 3 | 3 | 3 | 4 | 3 | 3 | 4 | 4 | 4 | 4 | 3 | 3 | |
| Becky | Expository | 1 | 1 | 2 | 1 | 3 | 2 | 3 | 2 | 2 | 3 | 2 | 3 | 1 | 2 | |
| Chris | Narrative | | | | | | | | | | | | | | | TS |
| Michelle | Persuasive | 4 | 5 | 3 | 4 | 3 | 3 | 3 | 4 | 5 | 4 | 5 | 5 | 3 | 3 | |
| . Laura | Persuasive | 2 | 2 | 3 | 2 | 2 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | |
| Jennifer | Narrative | 5 | 4 | 3 | 3 | 4 | 4 | 4 | 4 | 4 | 3 | 4 | 3 | 3 | 4 | |
| Tara | Persuasive | 4 | 3 | 4 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 3 | 3 | 5 | 4 | |
| Lueann | Expository | 4 | 4 | 4 | 3 | 3 | 4 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 3 | |
| . Michael | Expository | | | | | | | | | | | | | | | TS |
| . Kendra | Narrative | 3 | 2 | 2 | 2 | 3 | 3 | 3 | 2 | 3 | 3 | 2 | 2 | 4 | 3 | |
| Jeff | Narrative | 2 | 2 | 2 | 2 | 3 | 2 | 3 | 3 | 2 | 2 | 3 | 3 | 2 | 3 | |
| Ryan | Persuasive | 3 | 3 | 4 | 3 | 3 | 3 | 3 | 3 | 2 | 2 | 3 | 2 | 2 | 2 | |

SAMPLE 8

## Writing Assessment

School District
Grade 3

1991 Oregon Statewide Assessment

## Participation Report

| | District No. | District % | State No. | State % |
|---|---|---|---|---|
| Absent | 7 | 4 | 1143 | 3 |
| Exempted: Limited English Proficiency | 0 | 0 | 333 | 1 |
| Exempted: Special Education | 2 | 1 | 637 | 2 |

| | District No. | District % | State No. | State % |
|---|---|---|---|---|
| Modified: Limited English Proficiency | 0 | 0 | 255 | 1 |
| Modified: Special Education | 6 | 3 | 818 | 2 |
| Other | 0 | 0 | 114 | 0 |
| Number Tested | 164 | | 36340 | |

## Survey Questions

| 1. How much did you like writing on this topic? | District No. | District % | State No. | State % |
|---|---|---|---|---|
| Very much | 53 | 40 | 14369 | 45 |
| Somewhat | 49 | 37 | 8930 | 28 |
| Not much | 23 | 17 | 5631 | 18 |
| Not at all | 9 | 7 | 2681 | 8 |

| 2. How much do you usually like to write? | District No. | District % | State No. | State % |
|---|---|---|---|---|
| Very much | 43 | 33 | 12922 | 41 |
| Somewhat | 43 | 33 | 9511 | 30 |
| Not much | 38 | 29 | 7248 | 23 |
| Not at all | 5 | 4 | 1626 | 5 |

Mark all that are true

| 3. What kind of PLANNING did you do before you started to write your paper. Mark all that are true? | District No. | District % | State No. | State % |
|---|---|---|---|---|
| I made a list. | 40 | 30 | 11751 | 37 |
| I drew a picture. | 12 | 9 | 3110 | 10 |
| I just thought about my topic. | 96 | 72 | 20374 | 65 |
| I talked about my topic to someone else. | 16 | 12 | 3629 | 12 |
| I did something not listed here. | 17 | 13 | 6420 | 20 |
| I did not do ANY planning. | 14 | 10 | 3158 | 10 |

| 4. Did you write a rough draft? | District No. | District % | State No. | State % |
|---|---|---|---|---|
| Yes | 113 | 86 | 28754 | 92 |
| No | 11 | 8 | 926 | 3 |
| I do not know what you mean by a rough draft. | 8 | 6 | 1627 | 5 |

Mark all that are true:

| 5. What did you do to CHANGE your rough draft? | District No. | District % | State No. | State % |
|---|---|---|---|---|
| I read my paper over to myself. | 101 | 75 | 25306 | 80 |
| I took some stuff out. | 62 | 46 | 14513 | 46 |
| I put some stuff in. | 76 | 57 | 18921 | 60 |
| I moved some parts around. | 35 | 26 | 7266 | 23 |
| I left everything the way it was and just copied the paper over. | 19 | 14 | 4545 | 14 |
| I only wrote ONE draft of my paper. | 18 | 13 | 3353 | 11 |

Mark all that are true:

| 6. What things did you do to EDIT your paper? | District No. | District % | State No. | State % |
|---|---|---|---|---|
| I looked up some words in the dictionary. | 75 | 56 | 19773 | 63 |
| I read my paper over to see if I had left out any words. | 89 | 66 | 23099 | 74 |
| I fixed some mistakes. | 94 | 70 | 23104 | 74 |
| I tried to make the paper neater. | 63 | 47 | 17480 | 56 |
| I copied my paper over without changing anything. | 17 | 13 | 3577 | 11 |
| I only wrote ONE draft of my paper. | 16 | 12 | 3132 | 10 |

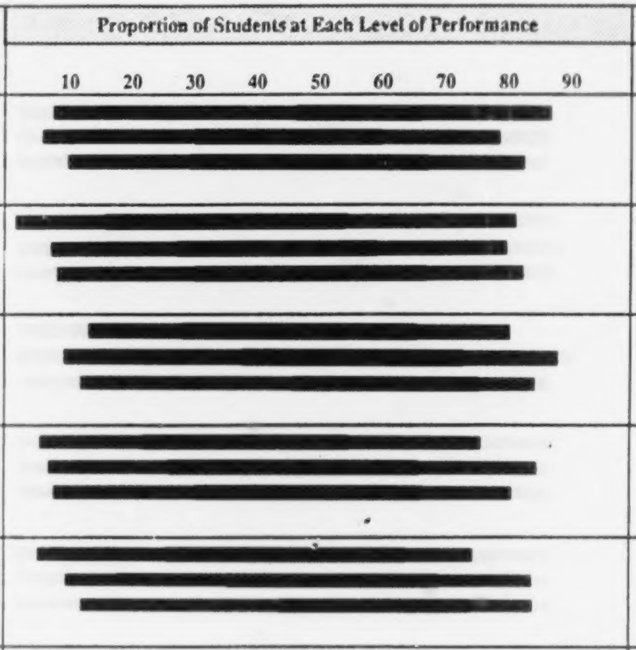
BEST COPY AVAILABLE

SAMPLE 9

## School Summary Report
## Writing Assessment

Elementary School
School District
Grade 03

**1991 Oregon Statewide Assessment**



| Mode: | 1 | 2 | 3 | 4 | 5 | No Score | |
|---|---|---|---|---|---|---|---|
| Descriptive | 6.7 | 40.0 | 20.0 | 20.0 | 0.0 | 13.3 | School |
| | 5.8 | 22.2 | 30.5 | 20.5 | 7.6 | 13.5 | District |
| | 8.3 | 23.7 | 32.4 | 17.5 | 4.5 | 13.7 | State |
| Narrative | 0.0 | 16.7 | 36.7 | 30.0 | 3.3 | 20.0 | School |
| | 6.3 | 20.8 | 32.7 | 20.9 | 8.8 | 10.3 | District |
| | 7.1 | 23.6 | 36.2 | 16.4 | 4.9 | 11.7 | State |
| Persuasive | 12.5 | 15.6 | 37.5 | 15.6 | 0.0 | 18.8 | School |
| | 8.5 | 28.4 | 36.4 | 15.1 | 1.8 | 9.9 | District |
| | 11.6 | 33.8 | 31.3 | 9.7 | 1.5 | 12.1 | State |
| Imaginative | 3.1 | 18.8 | 31.3 | 21.9 | 12.5 | 12.5 | School |
| | 5.2 | 19.9 | 39.6 | 21.4 | 7.1 | 6.7 | District |
| | 6.2 | 23.8 | 35.4 | 18.3 | 5.9 | 10.4 | State |
| Expository | 3.6 | 21.4 | 39.3 | 10.7 | 10.7 | 14.3 | School |
| | 8.4 | 26.5 | 34.6 | 16.3 | 4.6 | 9.5 | District |
| | 10.5 | 33.3 | 30.9 | 11.4 | 2.4 | 11.5 | State |

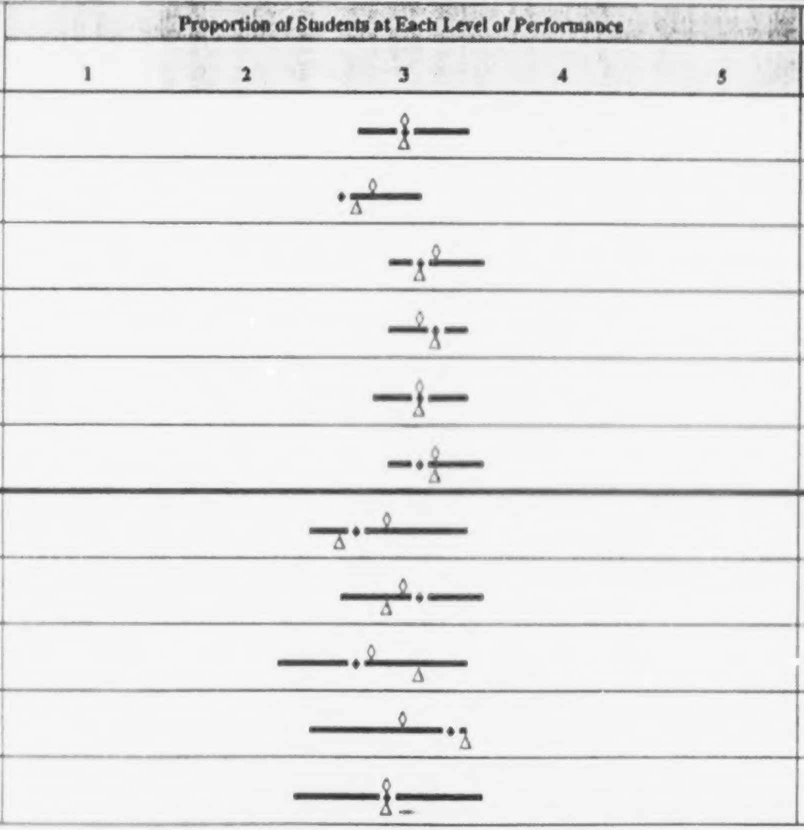
Percentage Scoring: 1 2 3 4 5

BEST COPY AVAILABLE

SAMPLE 10

**Writing Assessment**
**Comparison Group Report**
**Elementary School**
**School District**
**Grade 05**

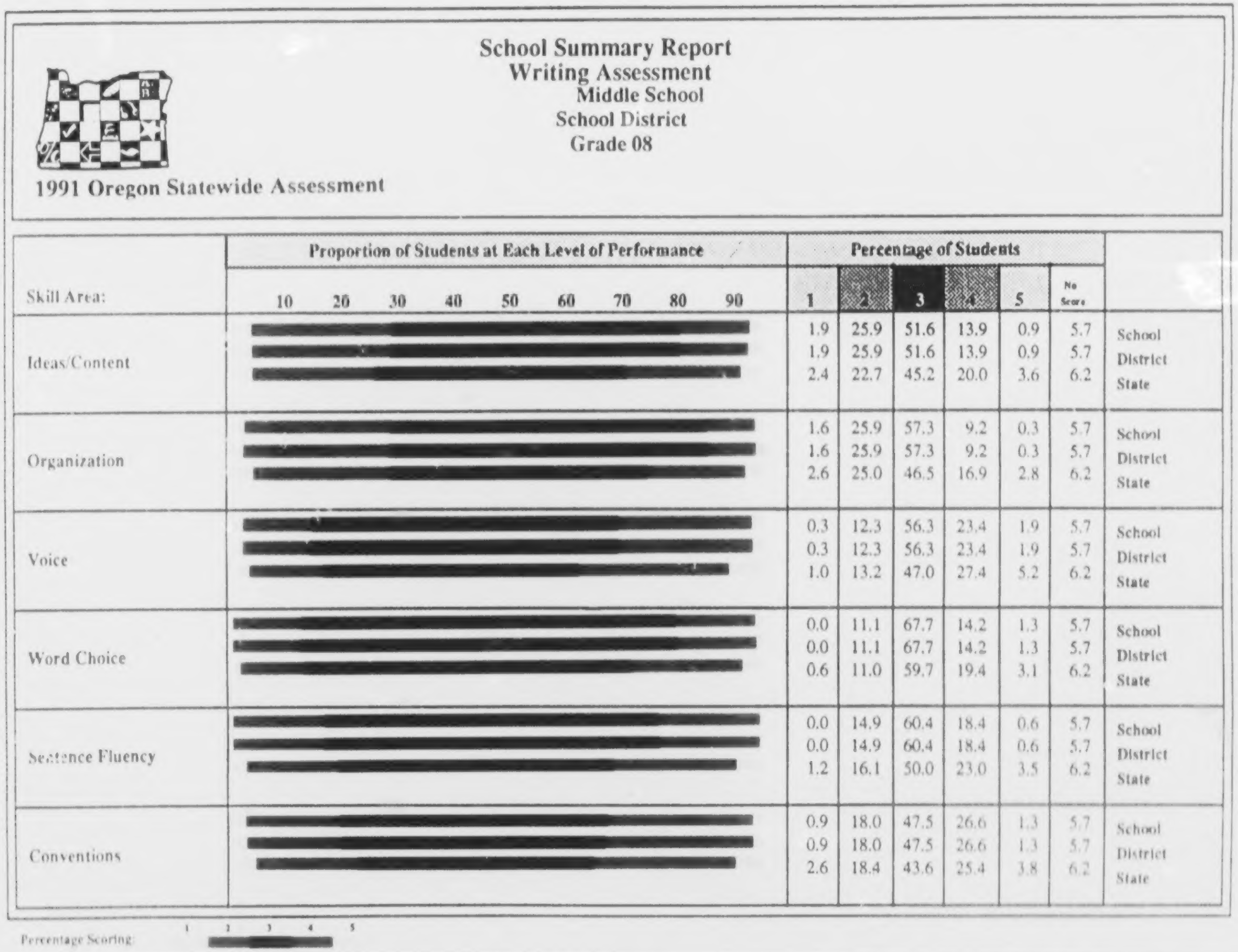
**1991 Oregon Statewide Assessment**



| | | Average Scores (♦) School | (Δ) District | (◊) State | (▬) Comparison Group |
|---|---|---|---|---|---|
| Trait | Ideas/Content | 3.0 | 3.0 | 3.0 | 2.7 - 3.4 |
| | Organization | 2.6 | 2.7 | 2.8 | 2.6 - 3.1 |
| | Voice | 3.1 | 3.1 | 3.2 | 2.9 - 3.5 |
| | Word Choice | 3.2 | 3.2 | 3.1 | 2.9 - 3.4 |
| | Sentence Fluency | 3.1 | 3.1 | 3.1 | 2.8 - 3.4 |
| | Conventions | 3.1 | 3.2 | 3.2 | 2.9 - 3.5 |
| Mode | Descriptive | 2.7 | 2.6 | 2.9 | 2.4 - 3.4 |
| | Narrative | 3.1 | 2.9 | 3.0 | 2.6 - 3.5 |
| | Persuasive | 2.7 | 3.1 | 2.8 | 2.2 - 3.4 |
| | Imaginative | 3.3 | 3.4 | 3.0 | 2.4 - 3.4 |
| | Expository | 2.9 | 2.9 | 2.9 | 2.3 - 3.5 |

BEST COPY AVAILABLE

## SAMPLE 11



### School Summary Report
### Writing Assessment

Middle School
School District
Grade 08

**1991 Oregon Statewide Assessment**

| Skill Area: | Proportion of Students at Each Level of Performance<br>10 20 30 40 50 60 70 80 90 | Percentage of Students<br>1 | 2 | 3 | 4 | 5 | No Score | |
|---|---|---|---|---|---|---|---|---|
| Ideas/Content | | 1.9 | 25.9 | 51.6 | 13.9 | 0.9 | 5.7 | School |
| | | 1.9 | 25.9 | 51.6 | 13.9 | 0.9 | 5.7 | District |
| | | 2.4 | 22.7 | 45.2 | 20.0 | 3.6 | 6.2 | State |
| Organization | | 1.6 | 25.9 | 57.3 | 9.2 | 0.3 | 5.7 | School |
| | | 1.6 | 25.9 | 57.3 | 9.2 | 0.3 | 5.7 | District |
| | | 2.6 | 25.0 | 46.5 | 16.9 | 2.8 | 6.2 | State |
| Voice | | 0.3 | 12.3 | 56.3 | 23.4 | 1.9 | 5.7 | School |
| | | 0.3 | 12.3 | 56.3 | 23.4 | 1.9 | 5.7 | District |
| | | 1.0 | 13.2 | 47.0 | 27.4 | 5.2 | 6.2 | State |
| Word Choice | | 0.0 | 11.1 | 67.7 | 14.2 | 1.3 | 5.7 | School |
| | | 0.0 | 11.1 | 67.7 | 14.2 | 1.3 | 5.7 | District |
| | | 0.6 | 11.0 | 59.7 | 19.4 | 3.1 | 6.2 | State |
| Sentence Fluency | | 0.0 | 14.9 | 60.4 | 18.4 | 0.6 | 5.7 | School |
| | | 0.0 | 14.9 | 60.4 | 18.4 | 0.6 | 5.7 | District |
| | | 1.2 | 16.1 | 50.0 | 23.0 | 3.5 | 6.2 | State |
| Conventions | | 0.9 | 18.0 | 47.5 | 26.6 | 1.3 | 5.7 | School |
| | | 0.9 | 18.0 | 47.5 | 26.6 | 1.3 | 5.7 | District |
| | | 2.6 | 18.4 | 43.6 | 25.4 | 3.8 | 6.2 | State |

Percentage Scoring: 1 2 3 4 5

BEST COPY AVAILABLE

# IX-2

## Schoolwide Portfolios

DONALD DAWS
METRO HIGH SCHOOL, CEDAR RAPIDS, IOWA

I am a mathematics teacher at Metro High School in Cedar Rapids, Iowa. I have also been a member of the building committee that has been working on implementing portfolios and alternative assessment since we began the process in the spring of 1990. At Metro we have defined portfolios as

> a record of learning that focuses on the student's work and her/his reflections on that work. Material is collected through a collaborative effort between the student and staff members and is indicative of progress toward our school's prescribed essential outcomes.

While portfolios in some schools focus on literacy or writing across the curriculum, ours is more inclusive. It includes all curriculum areas and all modes of learning from writing samples to audiocassettes to hands-on projects. Even though I have been asked to write this piece about our efforts, I want to stress that the progress we have made, and are continuing to make, is the result of the hard work and dedication of a number of individuals at Metro.

### Setting

Metro is the alternative high school for the Cedar Rapids Community School District. Our students come from all areas of the city, as well as from the surrounding metropolitan and rural areas. We serve approximately 600 students who range in age from 14 through 21. Our students are about equally divided between males and females. About 450 of our students are enrolled in our morning or afternoon programs and are working toward earning a high school diploma. The remaining students, about 150, are either enrolled in our Learning Center, where they are studying for their General Equivalency Diploma, or are temporarily inactive due to placement in rehabilitation or correctional facilities.

Our curriculum is designed to allow students to meet the prescribed graduation requirements of the district, although we often meet these requirements by using less traditional methods.

We have 37 teachers on the Metro faculty. Most of us serve as advisors for 10 to 25 students. Our role as advisors is to help students be successful at Metro. This includes assisting with educational planning; developing behavior-management strategies if necessary; and serving as a liaison between the school and parents/guardians, social workers, probation officers, and others when appropriate.

Teachers generally meet with students for six periods of an eight-period day, Monday through Thursday. Our administration makes an effort to limit class sizes to between 15 and 20 students for

---

Reprinted by permission of the NEA Professional Library from Donald Daws. "Schoolwide Portfolios." In *Student Portfolios*. NEA Teacher-to-Teacher Books. Washington, D.C.: National Education Association. Copyright 1993 by National Education Association.

classes taught by one teacher and between 20 and 30 students for team-taught classes.

Our students do not attend classes on Friday. We generally begin that day with a meeting of the entire faculty to discuss concerns and to coordinate Metro's program. During the rest of the day, we meet in a variety of committees and/or make home visits to our advisees.

## Basis for the Implementation of Portfolios

Metro has been a member of Ted Sizer's Coalition of Essential Schools since 1985 and a part of the National Education Association's School Renewal Network since 1989. Our connection with these organizations, the involvement of several of our faculty in graduate programs, and a continual growth in the percentage of our students enrolling in post-secondary institutions led to a number of us expressing a need to raise the academic expectations of our students. Further, we felt that if we increased our expectations, we also needed to develop new methods to document the students' progress in ways that expanded on our written narrative evaluations. Enter portfolios — a vehicle that would allow us to document authentic work in a central location and then use it as a tool in education planning and program evaluation.

## Chronology

**Spring 1990.** Each spring, schools in our district are required to develop a School Improvement Plan. This consists of a set of goals and subsequent strategies that the school staff plans to concentrate on the following year. To address our concerns with student expectations and additional documentation, the Metro staff determined that one of our committees for the implementation of the 1990/91 School Improvement Plan would focus on issues of curriculum, instruction, and assessment.

**1990/91 School Year.** Our committee decided, almost from our initial meeting in the fall, that portfolios and alternative assessments were two areas we would study and eventually recommend for adoption by the entire Metro staff. During the 1990/91 school year, our work in these areas focused primarily on developing a theoretical understanding of, and research basis for, our beliefs. Our group gathered, studied, and discussed numerous journal articles, research papers, and materials from the School Renewal Network that related to portfolios and alternative assessment. we found the work of Ted Sizer and the Coalition of Essential Schools as well as that of Joe Walters, William Spady, Grant Wiggins, and Dennie Palmer Wolf to be particularly helpful. Several of us also attended a state-level conference on restructuring. Finally, the district's Director of Research and Evaluation made a presentation to interested faculty members on the use of alternative assessments.

In the spring of 1991, we began a pilot program on portfolios. Several teachers labeled folders with student names and began collecting material from their classes for the portfolios. We also began talking individually with other faculty members to explain the portfolio concept and to encourage them to submit student work.

**1991/92 School Year.** In the fall of 1991, we decided it was time to present our committee's findings to the entire faculty. In October our group led a discussion to explain where it was in the process and to gather additional ideas and concerns regarding portfolios. Following this discussion, our committee developed a two-page handout that included our definition of portfolios, a rationale for their use, a proposed outline for mechanical procedures, and a description of the relationship between material selection and proposed essential outcomes for students (see Portfolio Reproducibles 1 and 2).

We also continued to talk with teachers on an informal basis. We felt this was necessary to alleviate individual concerns, answer questions regarding our rationale, and to encourage participation.

During the latter part of the year there seemed to be an unspoken shift in our philosophy: our committee focused its energies on increasing the level of teachers' submission of material and de-emphasized the need to connect these projects to one or more of the essential outcomes. The connection to essential outcomes had not been eliminated, but rather deferred in order to make the use of portfolios less cumbersome and/or threatening to beginners.

**1992/93 School Year.** In fall 1992 our committee continued to work informally to explain the process and rationale for portfolios. At this point, we began to see some benefits from our efforts. Staff participation increased and some of our students began to view their work in class in terms of inclusion in their portfolios. Some students were even asking their advisors for an opportunity to inspect their portfolios. Mary Vasey, our committee chairperson, decided to incorporate a portfolio review in the Preparing to Graduate Course for last-term seniors. This process assists seniors to realize what they have accomplished during their time at Metro.

Our committee also modified our portfolio response sheet (see Portfolio Reproducible 3). We did this in an effort to simplify the process for both students and teachers and to encourage a greater degree of reflection. We also suggested that teachers include a portfolio review as part of an advisor/advisee group meeting.

**Summer 1993.** During the summer of 1993, a number of staff members at Metro participated in a project designed to begin the development of authentic assessments. On the first day several members of the Curriculum Committee provided a brief overview of the rationale for the project. The purpose of this overview was to develop a common vocabulary, and to assist faculty members in connecting the Graduation Expectations with content material throughout the development of demonstrations (see Appendices A and B). Staff members then worked individually or in small groups to develop subject area, course, or unit demonstrations. We felt this degree of flexibility was necessary to allow individuals to work within their "comfort zones" during the initial stages of the project.

**1993/94 School Year.** The primary focus of our Curriculum Committee during the school year was to further delineate and clarify the Graduation Expectations (see Appendix C). At our annual staff retreat, we divided into four groups to discuss and refine each of the broad expectations. We plan to continue with these efforts during the upcoming school year.

In addition, there have been a number of staff members who have continued on an individual basis to incorporate demonstrations into classes, to increase portfolio submissions, and to explain to students how the Graduation Expectations are addressed in particular courses.

**Summer 1994.** This summer a group of five Metro teachers began a project to further integrate our efforts on Graduation Expectations, demonstrations, and portfolios. The materials we develop will be presented at our pre-service workshop in August for discussion by the entire Metro staff. Our discussions to date have focused on the need to provide concrete examples for possible portfolio submission (see Appendix D), and the development of an alternative portfolio response sheet that addresses the Graduation Expectations (see Appendix E).

We feel there is a need to provide staff time and direction in the use of portfolios with advisees. We plan to develop several activities which can be used by advisors to assist their advisees in connecting portfolios, Graduation Expectations, and course demonstrations. These activities will not only provide guidance, but will also focus attention on the importance of portfolios.

**Future Plans.** Our goal at this time is to continue to refine our current efforts, rather than begin new endeavors. We plan to continue to: develop demonstrations for content areas, courses, and units; unify curriculum, instruction and assessment; and integrate portfolios, Graduation Expectations, and demonstrations.

We also need to continue to assist our students in gaining an understanding of the connection among the Graduation Expectations, course content, and demonstrations.

It has been, and will probably continue to be, a slow process to gain schoolwide commitment and consensus. However, we are making progress in our efforts to develop a cohesive program for our students.

## Reflections

As I reflect on the last four years of portfolio implementation at Metro, five crucial areas surface. These are the time frame for implementation, the development of a theoretical basis, implementation methods, attention to personal concerns, and curricular changes.

### Time Frame for Implementation

As Barb Coates, one of our committee members, noted: "Portfolio implementation is a slow and evolutionary process." After four years, there are still a few teachers who either don't understand or don't accept a system of portfolios and alternative assessment. We are refining, and will continue to refine, our portfolio response sheets. We are continuing to develop methods that will allow for the effective and efficient use of portfolios by teachers, students, and advisors. Finally, and most importantly, we have not yet integrated portfolios and our essential outcomes in a cohesive manner.

I think when we started considering portfolios, we felt that we would have full implementation within a few years. Some of us have become somewhat discouraged with the slow rate of change. However, we should not have been surprised. Much of the research on the change process suggests that it takes from five to ten years for substantive change to occur.

### Development of a Theoretical Basis

I believe the process we used to develop our rationale for portfolios and alternative assessment was relatively solid. We started with a small group of individuals that was committed to the use of portfolios and willing to spend extra time researching the issue. Our group read and discussed a wide range of theoretical and practical material and then synthesized the information for discussion by the entire Metro faculty.

I also believe it was beneficial to distribute several key articles to all teachers and to provide time at a staff meeting for reading and discussing the material. Giving our faculty time to think about our proposals, ask questions, receive additional information, and then rethink the ideas seemed to increase acceptance and alleviate concerns about portfolios.

Our use of both formal and informal training methods proved useful. Our committee explained the key concepts and issues in a large group setting. This seemed to begin the development of a common set of understandings. The informal, individual discussions between our committee member and other teachers built on these understandings and served to ease some of the personal concerns.

### Implementation Methods

An idea we found useful, although it took us almost a year to move in this direction, was to separate the issues of portfolios and alternative assessment. Implementing both concepts as a unified package seemed to be more than some of our teachers could manage. As a result we seemed to be making little progress. When we concentrated on increasing participation and de-emphasized the need to relate the projects to the essential outcomes, teacher acceptance of portfolios seemed to increase. I don't believe this was due to a lack of harmony with the outcomes. Rather, some people seemed to need an opportunity to develop and incorporate routines involved in the portfolio process before they felt able to address and coordinate portfolio use with the outcomes. In fact, having a portfolio process in place has made it easier for us to begin focusing on the essential outcomes. However, we still need to address the extremely difficult process of linking these two areas.

This method is an example of what researcher Michael Fullan refers to as Ready-Fire-Aim. We developed a rationale for portfolios (got ready), focused on increasing teacher participation (fired), and now we need to connect portfolios and the essential outcomes (aim).

### Attention to Personal Concerns

During our initial discussions with the faculty, personal concerns were much more an issue than educational considerations. These concerns fell into two broad areas. The first related to both mechanical and procedural issues. Teachers asked questions like: "What will the portfolios look like?"; "Where will they be stored?"; "Who will be responsible for selecting, filing, and sorting student work?" These seemed much more urgent than ones like: "How will students benefit from portfolios?"

After a lengthy discussion, we arrived at tentative solutions to these questions. The folders would be stored in file cabinets located in an office near the library. One of our secretaries volunteered to label new folders and file student work. We decided that the teacher and student within a class would determine which materials would be included in the portfolios, and the advisor and advisee would be responsible for using the portfolio in educational planning.

The second, and related issue, dealt with time. Several of our teachers raised concerns regarding building in time for teachers and students to complete response sheets and providing advisors with the time to meet with advisees and work with them on a one-to-one basis.

Although our committee was not prepared for this reaction, we could have anticipated these concerns.

The literature on the change process stresses the personal nature of change. Until the personal concerns of the individuals affected by the change are addressed, little progress can be made on the educational issues.

### Curricular Changes

Probably the most profound effect on me personally has been the realization that portfolios, and their use in assessment, necessitate a change in the way I teach mathematics. If I truly believe in the process, then I can no longer have students move in a lock-step fashion through the textbook.

The other math teachers and I now intend to develop a set of key concepts in each area of mathematics. These will be used to develop projects that address the essential outcomes, either for use in our classes or in conjunction with teachers from other subject areas in interdisciplinary courses. Completed projects will be included in each student's portfolio for use by the student and his or her advisor. We will also use these projects to continually assess and update our mathematics program at Metro.

## Conclusion

At Metro we have made, and are continuing to make, progress regarding the efficient and effective use of portfolios and alternative assessment. Up to this point, our committee has done very well in some areas and not quite as well in others. I believe we could have foreseen and alleviated some of the difficulties we encountered if we had been more attuned to the literature on the change process. My suggestion to people about to begin portfolio implementation, or any innovation, would be to read, study, and incorporate the practical and theoretical findings of this body of knowledge as they start out on their journey.

PORTFOLIO REPRODUCIBLE 1

## Definition of Portfolio

A portfolio is a record of learning that focuses on the student's work and her/his reflection on that work. Material is collected through a collaborative effort between the student and staff members and is indicative of progress toward the essential outcomes.

**Rationale**
Portfolios can provide a vehicle to:

1. Assist students and staff members in assessing their progress toward acquiring the essential outcomes
2. Determine personal strengths, weaknesses, and preferences
3. Document the extent of students' willingness to take risks
4. Increase student self-esteem
5. Practice and emphasize reflection
6. Emphasize the importance of both product and process
7. Develop material for a senior seminar/exhibition that might serve as a portfolio for college and/or work
8. Assist with end-of-term evaluations
9. Develop both short- and long-term goals
10. Provide staff members with information to adjust course content and offerings to meet student needs

---

PORTFOLIO REPRODUCIBLE 2

## Mechanics of Portfolios

**Location**
Material will be stored in file boxes until and unless a more suitable method emerges. File boxes will be stored in Room 207.

**Acquisition of Materials**
The committee recommends that at least once per term the student and teacher select material for inclusion in the portfolio. The most appropriate time for this discussion, and the completion of a reflection sheet, is immediately following the completion of the project. Bonnie D. has agreed to file the material in student folders.

The committee recommends that at least once per term the student and advisor meet to review the portfolio. This conference can be used to develop the student's schedule for the next term and to establish goals in specific areas. Further, it can provide a method to highlight areas of growth and/or strength and emphasize the multidisciplinary aspects of central ideas.

## Essential Outcomes and the Selection of Material

Following a review of prior staff discussions and a sampling of the literature, the Curriculum/Student Outcomes Committee developed four preliminary essential outcomes.

Students, upon graduation from Metro High School will be:

**Information Literate**—Able to gather, organize, evaluate, synthesize, and communicate information.

**Self-Directed and Future-Oriented**—Able to set goals, to be aware of a wide range of possibilities and their consequences, and to meet and re-evaluate goals.

**Community Contributors**—Able to collaborate, participate, empathize, and communicate.

**Innovative/Creative Producers**

It is recommended that material selected for inclusion in the portfolio relate to at least one of the four essential outcomes.

PORTFOLIO REPRODUCIBLE 3

## Metro High School Portfolio Response Sheet

Name: ____________ Advisor: ____________ Date: ________

Class: ____________ Teacher ____________

Please describe your project. What did you do? Please be specific. ____________

Where did you get your ideas?
(Did you read about them? Ask someone? Use your own imagination?)
____________

What did you like most about doing this project? ____________

What frustrated you most? ____________

What do you wish you had done differently? ____________

Teacher comments: ____________

**DID YOU . . .**

write
observe
question
understand
envision
construct
put in your
  own words
plan
forecast
shape
think

compare
analyze
specify
survey
invent
photograph
share
identify
report
learn
support
sketch

organize
investigate
explain
describe
edit
relate
practice
create
eliminate
summarize
generalize
demonstrate

classify
contrast
interview
experiment
draw
present
express
solve
respond
brainstorm
draft
finish . . . ?

## APPENDIX A

### PLANNING BACKWARDS

↓

**GRADUATION EXPECTATIONS**
(general learner outcomes)

↓

**PROGRAM EXPECTATIONS**
(content areas)

↓

**COURSE/CLASS EXPECTATIONS**
(grade level or secondary course)

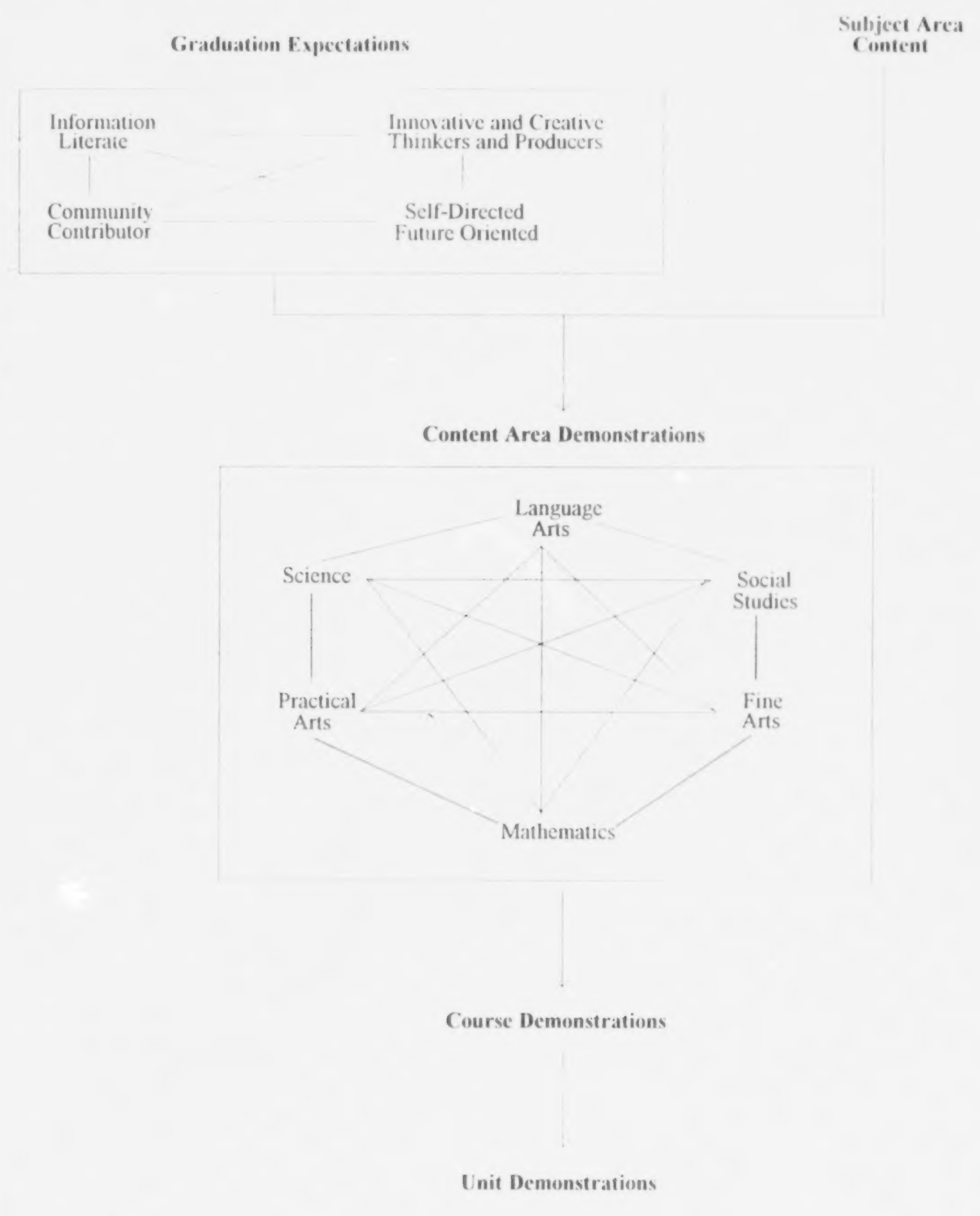
↓

**UNIT EXPECTATIONS**
(appropriate units to accomplish
course/class outcomes)

↓

**LESSON EXPECTATIONS**
(appropriate instruction)

## APPENDIX B


Graduation Expectations
Subject Area Content
Information Literate
Innovative and Creative Thinkers and Producers
Community Contributor
Self-Directed Future Oriented
Content Area Demonstrations
Language Arts
Science
Social Studies
Practical Arts
Fine Arts
Mathematics
Course Demonstrations
Unit Demonstrations

APPENDIX C

## Graduation Expectations for Students

**Information Literate**
- Identify, gather, process, organize, apply, evaluate, synthesize and document information
- Communicate information accurately and precisely for a variety of purposes and audiences
- Explain the process they used

**Self-Directed, Future-Oriented**
- Formulate a positive vision for themselves and their future
- Be flexible, and be aware of a wide range of possibilities and their consequences
- Set priorities
- Set, and meet or reevaluate realistic goals, persevering when the solution to a problem is not immediately apparent

**Community Contributors**
- Collaborate, cooperate, participate, empathize, and communicate in local and global communities
- Understand, accept, and respect differences

**Innovative and Creative Thinkers and Producers**
- Question
- Generalize, integrate, and transfer learning
- Solve problems
- Use the senses in appreciation and celebration of nature and art
- Show ingenuity, originality, insightfulness, creativity
- Show wonderment, curiosity, and enjoyment of learning

## APPENDIX D

## Possible Portfolio Items

**Information Literate**
- Papers, reports, and/or notes
- Bibliographies including materials from books, articles, interviews, videos, and/or CD ROMs
- Technical reading—following and interpreting directions
- Simulations
- Teacher assessments from class investigations
- Self-assessments, journals, reflections
- Use of technology
- Political cartoons
- Surveys
- Copies of letters asking for information
- List of all books in the library
- DRP, writing assessments, ACT, ITBS, ITED

**Self-Directed, Future-Oriented**
- Autobiography
- CHOICES, job shadowing, internships, and/or work evaluations
- Career planning, job-seeking skills inventory
- Independent projects
- Establishment of future goals
- Development of a plan for the completion of a project
- Career sequencing
- Use of technology
- Budgets
- Career/job knowledge
- Seminar planning class project and assessment project
- Letter to politician
- Interview person in a field of interest
- Job information
- Operating a small business
- Evaluation of goals

**Community Contributor**
- Record of activities that exhibit community involvement (awards, photographs, videos, articles and/or reports)
- Papers and/or stories which show respect for diversity and differences
- Brief response/reflection on a book or article relating to multicultural issues
- Part of a team
- Play productions
- Book report on other cultures
- Interview school, community and/or town leaders
- Interview community members

**Innovative and Creative Thinkers and Producers**
- Photographs, papers or other items documenting the completion of a creative project
- Problem-solving activities
- Evidence of the ability to analyze information, and to state and support (with evidence) a personal opinion
- Evidence of the ability to question, generalize, integrate, and transfer learning
- Use of technology
- Critiques, reviews, editorials
- Videotapes
- Theater/dance
- Lines and Shadows
- Independent projects
- Games designed to test knowledge of history
- General knowledge—reports, evaluations
- Ability to work on several things at once
- Seek answers and knowledge from others, experts
- Daycare checklist

APPENDIX E

## Portfolio Reflection Sheet

Name: ______________________ Advisor: ______________________ Date: __________
Class: ______________________ Teacher ______________________

1. **Stage of work:**
   - ___ Original idea/sketch, rough draft.
   - ___ Second draft, almost the way I want it.
   - ___ Finished project.

2. **Choose one or two of the following Metro Graduation Expectations, and explain why your project addresses the Expectation(s):**
   - ___ Information Literate because:
   - ___ Self-Directed, Future Oriented because:
   - ___ Community Contributor because:
   - ___ Innovative and Creative Thinkers and Producers because:

3. **This item was included because:**
   - ___ The teacher wants me to include it.
   - ___ It was a challenge for me.
   - ___ It represents my ability at this time.
   - ___ It was good.
   - ___ I liked it.
   - ___ Other. ____________________

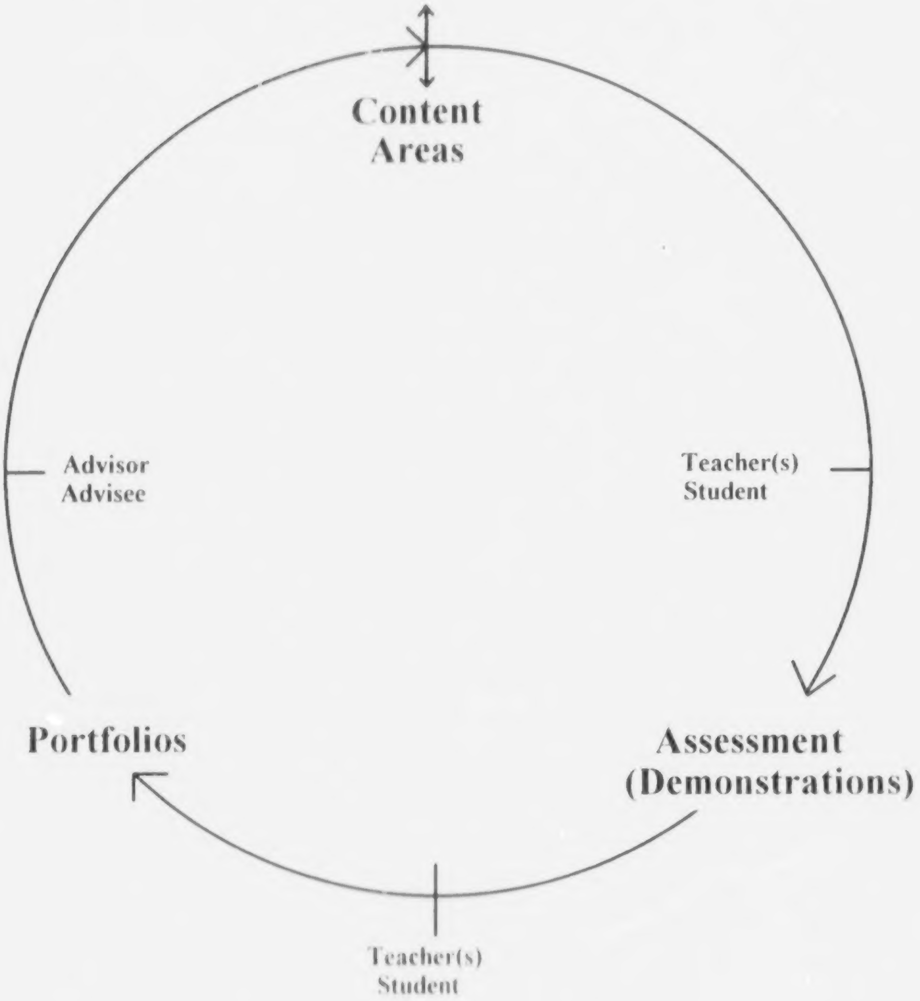
4. **Teacher/Student Comments:**
   Talk through process of project (goals, expectations, assessments . . .) with the teacher. Please write the initials which best describe how you felt about progress on this project.

NP - no problem SC - some challenge for me NW - needs work

| Student | Teacher | |
|---|---|---|
| ____ | ____ | getting started |
| ____ | ____ | mechanics and/or technical skills |
| ____ | ____ | following directions |
| ____ | ____ | researching |
| ____ | ____ | organizing the parts |
| ____ | ____ | finishing the project |
| ____ | ____ | self-directed |
| ____ | ____ | working with others |
| ____ | ____ | other (explain) ____________________ |

APPENDIX F

## Graduation Expectations


Content
Areas
Advisor
Advisee
Teacher(s)
Student
Portfolios
Assessment
(Demonstrations)
Teacher(s)
Student

# IX-3

## Assessing What We Value: Snapshots from Waialae's Work in Progress

KATHLEEN U. BUSICK
PACIFIC REGIONAL EDUCATIONAL LABORATORY

What does creative problem solving look like in the work of very young children? Does every task have some kind of problem solving embedded in it? What are our standards? How do they connect with emerging state and national standards? Do we need to develop rubrics for our creative problem-solving criteria that describe what creative problem solving looks like for children at different stages of development? How do we deal with the extremes of student self-assessment—those who are very hard on themselves and cannot see the positive qualities in their work, and those who see their work as more fully developed than it actually is? How can we report to parents in ways that are much richer and more clearly related to what we value while we are required to use a statewide report card?

These and many other questions are part of the conversations underway at Waialae Elementary School as its school community works toward a vision for student success. Assessment is an integral and essential part of those discussions.

### Context and Goals for Students

Waialae Elementary is one of Hawaii's School Community-based Management (SCBM) sites. Located in suburban Honolulu, Waialae Elementary School has about 500 students, 19 full-time classroom teachers, and one administrator. Three years into restructuring curriculum, instruction, and assessment to improve student learning, teachers, students, parents, administrators, support staff, and community members have worked toward sharing a vision for success that includes the following student goals:

Waialae's students shall become:

- Creative problem solvers
- Self-confident risk takers
- Well-rounded: capable in multiple dimensions
- Collaborative and socially responsible to others and the world

This vision arose from the school community's belief that the role of education today is to provide

---

Copyright © 1994 Pacific Region Educational Laboratory. Reprinted with permission.

children with the understandings, skills, and attitudes necessary to interpret and contribute to an ever-changing and complex world. Our children will face choices and decisions that are not within the realm of today's realities. Traditional curricula and assessment do not adequately help children develop the knowledge, competencies, and dispositions they will need to effectively and creatively deal with life's challenges.

The role of assessment within this restructuring school is evolving in response to the vision and the growing experience of teachers. The process is a recursive one in which the classroom implementation of assessment strategies, including vision-based portfolios, student self-reflection and self-assessment, performance tasks, an integrated curriculum, and active hands-on learning, results in more questions, serious review, reflection, and the refinement of criteria, descriptors, integrated units, reporting forms, and teaching practices. We have come to understand that assessing significant goals requires us to continually reflect, frequently examine student work across grade levels, periodically adjust our teaching strategies and curriculum, and to occasionally step back and reconsider earlier decisions. There have been no easy shortcuts.

## Expanded Definitions/Performance Criteria

For the Waialae community, building on a complex vision has meant defining the elements of the vision in operational terms, getting students involved in making the vision real in their own terms, and using samples of student work to further refine the multidimensional assessment system. Each year has seen a deepening of the questions asked and the answers constructed by teachers, students, parents, and community members.

For us the process of developing criteria that describe creative problem solving is also the process of clarifying our goals for students. Performance criteria *are* expanded definitions. Figure 1 shows the current draft of Waialae's Standards and Indicators for Creative Problem Solving. Defining and describing creative problem solving has been an ongoing process. A first draft was developed by a task force of teachers who began by sorting samples of student responses to the same mathematical problem.[1] Individually and then in small groups, teachers placed student responses in three piles. One pile contained responses that were considered excellent examples of problem solving; another pile contained responses that showed developing problem solving skills—work with strengths and weaknesses; and a third pile contained responses that indicated beginning problem solving. Working in small groups, the teachers compared their placement of the responses and recorded the evaluative terms used to characterize the responses in each pile, and the evidence cited for the judgments made. Where there were differences in placement, clarification of reasoning and further examination of student work resulted in growing consensus about the qualities of different levels of response.

We then examined a variety of existing criteria and scoring guides to expand the possibilities and check for clearer language. The first round included discussions of the differences between analytical and holistic scoring and the appropriate uses of each for various purposes. Although mathematical problem solving began the development of criteria, the vision is not limited to mathematics; it also includes student creativity. Both stretch and challenge the school community to clarify the language of the criteria and the focus of the curriculum. Parents, too, are contributing to the refinement of criteria as they seek to apply them to student- and teacher-selected work in portfolios.

As work continues, we often come back to this process—gathering samples of student work, looking for evidence of the qualities we have struggled to define together, and modifying the criteria and descriptions as better and clearer understanding develops.

The language of assessment is a constant concern. We have variously labeled items as criteria, standards, and dimensions; we have talked about indicators, descriptors, and more. We have struggled to communicate clearly with all who want to take part in Waialae's growth. And we have also confronted the issue of how we view assessment—as measurement or as portrayal of student learning. Waialae has committed itself to gathering rich and detailed information about student learning. The terminology and mind set of "measurement" seems confining. At the same time, we want the portrayal of learning be accurate and truthful.

There have been three drafts of standards and descriptors for Creative Problem Solving in the past year alone. Additions and modifications are still taking place as teachers involve their students in brainstorming and contributing to the language of portrayal. Figure 2 shows 3rd graders' contribution that resulted from a discussion about jobs of the future and the things they thought they needed to be prepared for that future.

---

[1] Exercise developed by Judith A. Arter, Northwest Regional Educational Laboratory.

FIGURE 1

## WAIALAE SCHOOL PRELIMINARY DRAFT OF STANDARDS AND INDICATORS FOR CREATIVE PROBLEM SOLVING

### Draft #3

**Definition of Creative Problem Solving**

Creative problem solving includes the ability to understand the problem, build solutions, and follow through and produce an outcome which addresses or solves the problem. It includes an element of creativity, which requires that children extend beyond their own limit in the problem solving process or outcome, and involves the ability to see a multitude of possibilities. Creative problem solving is developed through continuous opportunities to perform authentic tasks and involves children being able to construct meaning or find connections within the larger context of real world situations/problems.

**Standard #1: Understands Problem**

- Distinguishes between relevant and irrelevant information and factors in the problem
- Links experience and prior knowledge to problem
- States the problem in a question or series of questions
- Reflects on possible multiple approaches to arrive at outcomes and/or possible multiple outcomes for the problem
- Relates problem to real world situations and contexts
- Makes meaning of the problem in personal, social or community sense
- Interprets problems from other points of view and multiple perspectives

**Standard #2: Building Solutions and Following Through**

*Making a Plan*

- Draws on and applies past knowledge
- Identifies information and resources
- Evaluates multiple approaches
- Selects appropriate, feasible, efficient, simplest solution strategy
- Describes plan of action

*Follow Through*

- Accesses and gathers appropriate materials and data
- Takes risks by trying ideas that are new or original for the child
- Evaluates appropriateness of strategy
- Adjusts strategy based on evaluation
- Explains and describes what worked and did not work during process to create final product

**Standard #3: Outcomes**

*Originality*

- Product is a stretch of child's limits
- Product is unique for child
- Uses alternative resources/materials

*Quality*

- Product is clear and coherent
- Product is accurate
- Product is refined and polished
- Product is organized
- Demonstrates effort
- Demonstrates persistence

*Appropriateness*

- Product or process is related or connected to problem or task
- Reflection
- Evaluates product and makes suggestions for improvement

**Standard #4: Creativity**

*Flexibility*

- Generates multiple approaches

*Transfer in Thinking*

- Applies approaches to problems to new solutions and contexts
- Communicates verbally and nonverbally about application of concepts to new or real world situation

*Originality*

- Demonstrates divergent thoughts and approaches which are different from past thoughts and practices for that child
- Demonstrates divergent thoughts and approaches which are unique for all
- Reflects on the uniqueness of thoughts and approaches

BEST COPY AVAILABLE

## FIGURE 2
## STUDENT CONTRIBUTIONS TO CRITERIA

### Waialae's Vision for Its Children
### Grade 3 Vision Setting

**Creative Problem Solver**

1. Help each other
2. If you are in trouble, need to know who to ask or solve it yourself
3. Solve problems in many ways
4. Help people when they're hurt or stuck
5. Be alert
6. Need to know how to use things like controls and computers
7. Have to know what you are doing
8. Listen carefully and follow directions
9. Work hard
10. Get a good education (basic skills)
11. Must be intelligent in many ways
12. Do your homework
13. Don't fight but form a consensus
14. Stay in school

**Self-Confident Risk Taker**

1. Be confident in what you're doing
2. Be brave in many ways
3. Use your time wisely
4. Learn a foreign language
5. Know how to survive
6. Need to know how to use and make food, water, and equipment
7. Know how to use the computer
8. Must be strong
9. Make sure everything is working right
10. Know how to use the dictionary and encyclopedia
11. Be happy
12. Feel good about yourself
13. Tell the truth
14. Don't use drugs or drink
15. Take a bite out of crime
16. Must have a clear mind
17. Do your best in everything
18. Believe in yourself
19. Be careful
20. Be alert
21. Don't play with dangerous objects
22. Must be intelligent in many ways
23. Don't think you're junk because you really might get better
24. Control your temper, be mellow

**Well-Rounded: Capable in Multiple Dimensions**

1. Be flexible in body and mind
2. Be very fast
3. Be healthy in body and mind
4. Can't be lazy
5. Study hard and get a good job
6. Try not to act up
7. Don't get so hyper
8. Don't brag or show off
9. Don't act too smart
10. Be careful in many ways
11. Listen to the teacher
12. Don't answer back
13. Be active
14. Must be intelligent in many ways
15. Be honest
16. Be considerate

**Collaborative and Socially Responsible to Others and World**

1. Cooperation
2. Teamwork
3. Need to be friendly
4. Be a good citizen
5. Help the team
6. Be kind
7. Care for the environment
8. Care for each other
9. Treat other people how you want to be treated
10. Tell the truth
11. Don't cheat
12. Use common sense
13. Feel good about yourself
14. Don't use drugs or drink
15. Take a bite out of crime
16. Must have a clear mind
17. Don't answer back
18. Help each other
19. Don't push other people around
20. Don't be bossy
21. First come, first served unless first person says it's okay to switch
22. Don't tease other people
23. Don't say anyone is dumb or call names because they might be better than you
24. Judge people from the inside
25. Don't start fires
26. Fix things that are broken or dangerous
27. Recycle and/or reuse items

BEST COPY AVAILABLE

FIGURE 3
A-1'S CREATIVE PROBLEM SOLVING CRITERIA

**Identify the Problem**
- Figure our that there is a problem
- Know what the problem is
- Find out what the problem is

**Understand the Problem**
- Find out clues about the problem to help me better understand the situation

**The Process of Solving the Problem**
- Form a list of possible solutions
- Think of different solutions
- Find the "best" plan
- Pick the "best" solution
- Eliminate the solutions/plans that don't work

**Evaluate/Assess the Situation:**
- Is it working?"
- Make a plan of action

**Evaluate/Assess the Situation:**
- Is it working?"
- Am I headed toward my outcomes?"
- If it's not working, go back to my list of solutions
- Carry out the plan of action

**Evaluate/Assess the Situation:**
- Is it working?"
- Am I headed toward my outcomes?"
- If it's not working, go back to my list of solutions

**Evaluate the Solution**
- See if my solution worked

Fourth graders, too, have been working to describe the key qualities of creative problem solving. Figure 3 shows their list of problem-solving criteria. Throughout the school, at various levels and in different ways, teachers and students are entering into extended conversations about the vision that guides their work.

## Portfolio Processes

Students' selection of work for their portfolios, their reflections, and self-assessment are key parts of Waialae's assessment activities. When asked to select examples of their work as creative problem solvers, some children chose practical and functional problems like "I didn't know how to cut the pieces." When asked how they solved the problem, "I asked Mom" was a common response. It was hard for teachers who worked hard to create tasks with opportunities for more significant problem solving to accept this level of problem solving. We were forced to go back to our definition of creative problem solving and remind ourselves that overruling a student's choice means losing the student ownership we are striving for. We also struggled with the worry that parents would not get a full picture of their children's learning if all portfolio selections were made by the student. Some teachers, especially at the lower grades, have decided that they will reserve the right to include picks of their own in the student portfolios to give a fuller picture of student learning. Others have been communicating with parents about the process that students are going through and the reasons for collection, selection, reflection, and student self-assessment.

## Strategies Resulting in Rich Information About Creative Problem Solving

### Classroom Projects and Tasks

Figure 4 is an example of a form developed by Waialae's kindergarten teacher to encourage students to reflect and respond in pictures to the problems embedded in the stories they are hearing and reading. Students have also been asked to propose two alternative solutions to the story's central problem and to select the one they think is the best. As creative problem solvers themselves, teachers have invited upper elementary students to come in and record kindergartners' descriptions of their drawings. Students are creating webs and concept maps early in their units and including those and later webs in their portfolios. Free writing and dialogue journals are also a part of the process for some classes. Teachers are exploring this year—trying out a variety of strategies and borrowing good ideas from each other.

### Student Reflection

Reflection by students emerged as an issue after the first round of student portfolio selections. Teachers studied sample prompts from other schools, read articles, and communicated with educators from other schools and districts in preparation for drafting reflection prompts for their students. At first, teachers tried using common prompts across levels and kinds of tasks. When all students made their first selections from work folders for their portfolios, they were asked to describe the selections and the reasons for their choices. Initial prompts focused on reasons for selecting most satisfying work, unsatisfying work, best work, and free picks. The results prompted serious concern. Initial reflections seemed shallow for many students. We discussed the growth of self-reflection and the need for modeling by teachers. Pri-

BEST COPY AVAILABLE

FIGURE 4

## CREATIVE PROBLEM SOLVING

Name

1. Draw the problem.

2. Draw your solution.

3. Circle your best solution.

4. What do you think of your solution?

mary teachers were especially concerned about how to capture student reflections. Figure 5 shows a form used in 1st grade to prompt students to think about their learning.

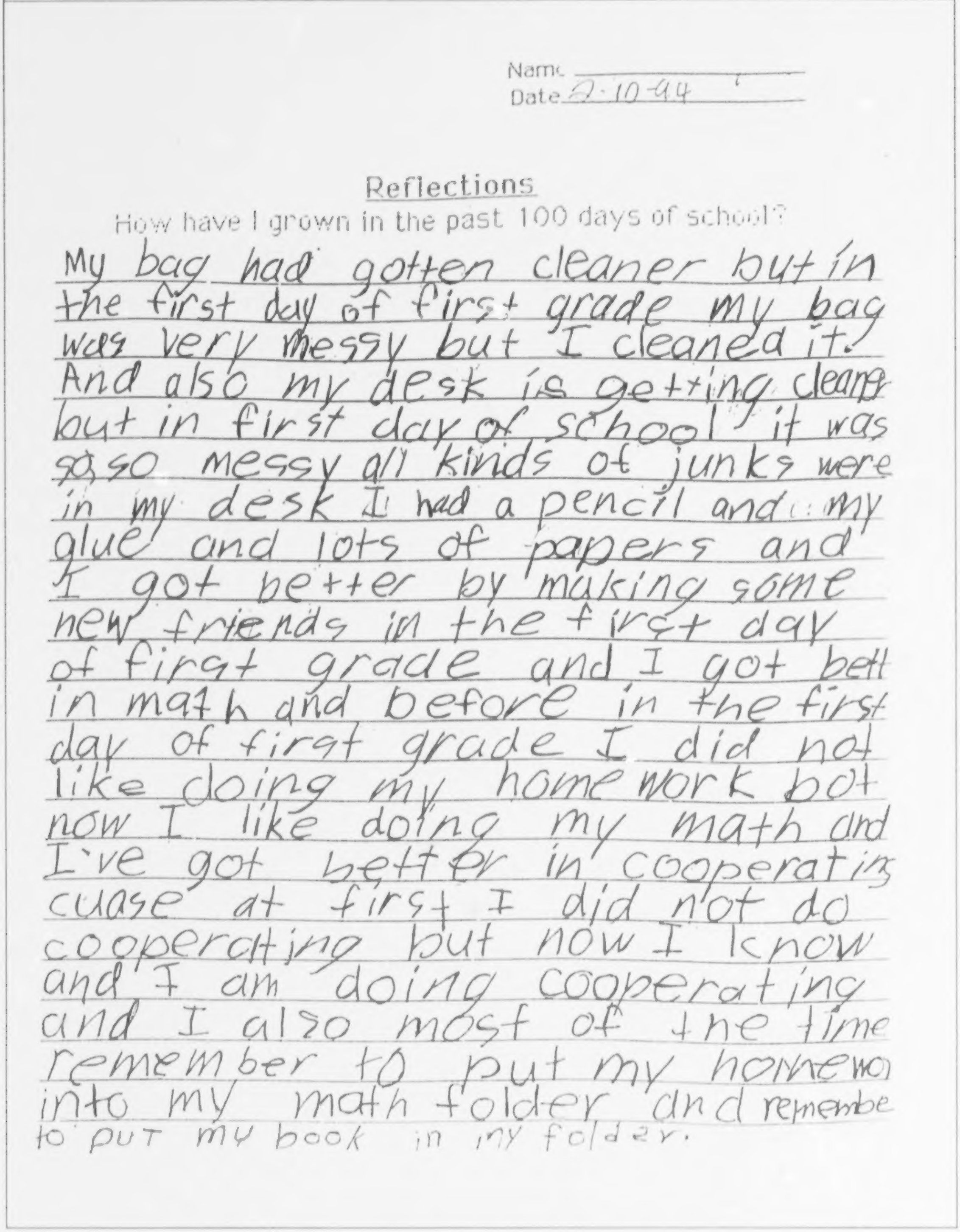
## Using Assessment Results for Improvement

Using draft performance criteria to analyze student responses to creative problem solving tasks has prompted teachers to examine their planning of integrated units. At times it seemed necessary to force problem solving into unit activities or to contort the units to accommodate creative problem solving. Some teachers worried about trying to simultaneously develop a portfolio culture and implement jointly planned units. Assessment in the form of portfolios for creative problem solving seemed like an addition rather than an integral part of the units. One modification in planning that grew out of this year's struggles has been to directly identify opportunities for assessment within unit activities. Teachers are identifying the standards addressed in their units as they prepare for the 1994–95 school year.

## Critical Questions

Much work is in progress; much remains. We began with questions about how to better assess student work in the midst of significant change in curriculum and instruction. Procedural and information questions dominated the first year. Management questions took over as the system began to develop. This year teachers, parents, and students are asking deeper questions than last year. Concerns about management and information are still there, but the effects of the changes are more often the focus of discussion. Figure 6 displays a sampling of Waialae's current questions. Answers are being crafted together and individually. Answers, too, often lead to more questions. The journey is well underway, the vision is getting clearer, but bumps in the road are normal, and the tears and cheers are beginning to balance out. Waialae School welcomes ideas and suggestions. Like their students, Waialae's teachers, administrator, parents, and community members are growing together.

FIGURE 5
"REFLECTIONS" FORM USED IN 1ST GRADE



Name
Date 2-10-94

Reflections

How have I grown in the past 100 days of school?

My bag had gotten cleaner but in
the first day of first grade my bag
was very messy but I cleaned it.
And also my desk is getting cleaner
but in first day of school it was
so, so messy all kinds of junks were
in my desk I had a pencil and my
glue and lots of papers and
I got better by making some
new friends in the first day
of first grade and I got bett
in math and before in the first
day of first grade I did not
like doing my home work bot
now I like doing my math and
I've got better in cooperating
cuase at first I did n'ot do
cooperating but now I know
and I am doing cooperating
and I also most of the time
remember to put my homewo
into my math folder and remembe
to put my book in my folder.

BEST COPY AVAILABLE

FIGURE 6

## TEACHERS' CHOICES AND QUESTIONS AS THE SYSTEM EVOLVES

As Waialae School journeys toward a vision of students who are creative problem solvers, self-confident risk takers, collaborative, and capable in multiple dimensions, teachers have become deeply reflective about the choices they are making with the school community, and the impact of those choices on students, parents—all the partners in this long-term process. Below is just a sampling of the kinds of questions that have come from teachers. They seek to construct answers that are consistent with the school community's vision.

| Choices | Related Questions |
|---|---|
| **Develop a vision-based portfolio** | • What models can we use?<br>• Will others understand what they see in student work?<br>• Have we sufficiently defined each component of our vision?<br>• Do our standards adequately describe the qualities that represent each component of our vision?<br>• Will this involve larger changes in our curriculum and instruction? |
| **Begin with creative problem solving** | • What IS "creative problem solving"?<br>• Do all samples in this part of the portfolio need to have some element of "creativity"?<br>• Is it "creative" if the work shows something that's unique for the child even though it might not be "unique" in and of itself?<br>• Does this mean that we have to focus on problem solving in most of our instruction?<br>• Does every task involve some form of problem solving? How broadly should we define this component?<br>• Will we end up contorting our lessons to be sure we fit in opportunities for problem solving? Is this what we want?<br>• How clear is the language of our creative problem solving standard for our students? for parents? for others who want to participate in the conversations about what our students can do? |
| **Involve students in critical decisions about their portfolios** | • How do students learn to value their work?<br>• How do we help them make choices that clearly document what they know and are able to do in relation to creative problem solving?<br>• What's the collection for? How long do they collect before they select?<br>• Is the time frame different for different ages?<br>• How long does it take to build a portfolio culture in which students value their work, understand the criteria, and make "good" selections?<br>• If we model and discuss in advance, does this compromise the portrait? Does it leave us without a baseline? And is that assessment purpose more important than other purposes for the portfolio?<br>• How does our role change? |
| **Use portfolio for assessment: to gather rich information about students' creative problem solving** | • How do we build in opportunities for gathering information?<br>• What kinds of things go in the portfolio for very young students?<br>• Will their limited language skills mask their learning?<br>• Do we need to agree on forms to use?<br>• Do we need to use common prompts for student reflection?<br>• How can we physically get the time to do student interviews in the midst of instruction?<br>• What kinds of problems are students identifying? Do we need to help them distinguish between significant and relatively trivial problems? |

(continued)

BEST COPY AVAILABLE

FIGURE 6—*(continued)*

## TEACHERS' CHOICES AND QUESTIONS AS THE SYSTEM EVOLVES

| | |
|---|---|
| **Begin looking at using the evolving portfolio system for public accounting in addition to instructional decision making and student assessment** | • What's the role of the advisory group?<br>• Will they find us lacking when they examine initial student portfolios?<br>• Will they understand that we're still working on determining the standards?<br>• Why can't we build the portfolio culture first and then after we've gotten clearer and more secure invite others in to look over our shoulders?<br>• Will they understand how much our vision involves change—in our teaching, our relationship with students and parents, as well as our assessments?<br>• What will be the consequences if they look before we're ready? |
| **Implement integrated curriculum while developing the multi-dimensional assessment system disciplines?** | • How do we fit our standards into the curriculum units we're developing?<br>• Can we undertake this many changes simultaneously?<br>• What's not worth hanging onto?<br>• Should we gather information about the vision elements only, or should we also try to assess traditional<br>• How do we pull out useful information about reading, math, etc., when we are teaching thematically?<br>• How does assessment connect to out units? |
| **Involve students in reflecting about** | • What is self-reflection? How can we teach students to be reflective?<br>• When should they reflect?<br>• Why should they reflect?<br>• Are there ways for very young students their work to reflect?<br>• Do we need to have our students reflect on every piece they select for their portfolio?<br>• If we use the same prompts over and over agair to be sure that we have similar data for assessment, won't they become just another requirement that really doesn't give us insights into students' learning?<br>• What good are prompts for self-reflection?<br>• Do we assess student reflections?<br>• What if student reflections are shallow? |
| **Involving students in assessing the quality of their work** | • What's the difference between self-reflection and self-assessment?<br>• What criteria do they use for self-assessment?<br>• How do we use the self-assessment information?<br>• What if their judgment doesn't match ours?<br>• Should we invite parents into the process too? |

BEST COPY AVAILABLE

# IX-4.

# The Senior Project: South Medford High School

NANCEY OLSON
NORTHWEST REGIONAL EDUCATIONAL LABORATORY

This paper illustrates how school districts can use assessment strategies that are more closely aligned with the important outcomes of schooling. For years teachers, administrators, parents and students have complained about the inadequacy of the testing programs most commonly used in school districts. The most frequently used data come from the annual administration of norm-referenced standardized tests. Research indicates that this is an inadequate measure of the school curriculum, especially the broad learner outcomes that are more difficult to measure. These test results provide limited information for program improvement and usually only in the basic skill areas. In addition, even though the tests are not intended to provide adequate information about individual student growth, they are often the only source of feedback that parents and students have of a consistently administered measure.

Because of this dilemma, more and more individual schools and some districts are turning to alternative and/or additional measurement strategies that provide them with more closely aligned measures of student performance. One approach, and the focus of this paper, is the Senior Project, which is viewed as a culminating experience of considerable significance to students and staff. At South Medford High School in Medford, Oregon, the staff has designed and implemented Senior Projects as one way of assessing student outcomes and a way of meeting Oregon State Standards that require individual student assessment and instructional program evaluation. This graduation requirement for all South Medford seniors requires in-depth gathering of information for a research paper, the application of skills and knowledge in the production of a project related to the paper, and a presentation of their work to a panel of judges consisting of teachers and community members knowledgeable in the area of the project.

## The Senior Project: Overview

The rest of this paper presents a case study based on South Medford High School's Senior Project. The four parts of the paper are as follows:

**Part 1—Curriculum Alignment.** This section describes how the Senior Project, as an integrated assessment model, is aligned with the Essential Learning Skills and the Common Curriculum Goals as outlined by the Oregon Department of Education.

---

Adapted from Nancey Olson, *Curriculum Aligned Products: The Senior Project*. Portland, Oreg.: Northwest Regional Educational Laboratory, 1993. Reprinted with permission from South Medford High School.

**Part 2—Senior Project Portfolio.** This section provides a description of the three parts of the Senior Project and copies of the documents contained in the Senior Project portfolios. These forms are used: (a) by the students to guide their planning and self-evaluation; (b) by the readers to critique and then grade the research papers; © by the judges to rate the presentations; and (d) by parents to verify the students' work on the actual projects.

**Part 3—Examples of Data Displays.** This section illustrates how data could be presented if the school chose to aggregate the individual student data to the building level.

**Part 4—Issues and Concerns.** This section presents management and record keeping issues as identified by the ODE in the paper *English Language Arts, Report 8, Integrated Assessment Model: A Project-Based Approach* (September 1988).

## Part 1: Curriculum Alignment

The Oregon Department of Education (ODE), Division of Curriculum and School Improvement, has issued curriculum guidelines containing goals that districts are to integrate into the development of their written curriculum documents. These guidelines are called the Essential Learning Skills, which are those goals that all teachers are responsible for integrating across all subject areas, and the Common Core Curriculum, which are those goals that are subject-area specific.

As stated in the ODE paper *English Language Arts, Report 8*:

> Curriculum goals which have been focused on factual information within a subject area will shift to broader conceptual frameworks open to inquiry and experimentation. Through technology, facts are readily available. It is the relationships and syntheses of facts that students need in order to understand and evaluate the importance of information. The Oregon Common Curriculum Goal documents reflect the shift of emphasis from isolated elements of knowledge (facts) to integrated conceptual learning; that is they give as much attention to how one learns as to what one knows.

The goals for this restructured curriculum state that students will be:

- Independent learners—curious, interested, life-long learners
- Resourceful, creative, inventive, problem solvers, decision makers
- Achievers, completers, goal-focused
- Competent, knowledgeable, in-depth experts, confident of what they know, aware that they don't know all
- Productive, engaged, at-task, able to sustain interest
- Literate, able to communicate and to receive communication (1988, pp. 2-3).

The ODE states, "The philosophy behind the Essential Learning Skills and the Common Curriculum Goals is that schools should prepare students to function effectively beyond high school because of their education, not necessarily with the education they received in 12 years" (1988, p. 4). The project assessment model requires evidence of how the student learns, how well the student manages the learning process, and what the student knows beyond what is regularly tested through a single, limited measure of knowledge and skills.

Many Essential Learning Skills are difficult to measure through typical pencil-and-paper, multiple-choice achievement tests. The Senior Project model used by South Medford High School is aligned with the Essential Learning Skills (1988, p. 5) shown in Figure 1.

## Part 2: Senior Project Portfolio

The Senior Project at South Medford begins second semester with an explanation to the students of the requirements of the project, and each student responds by writing a letter of intent to the English teacher who will guide the student through the project. The letter of intent states what topic the student initially plans to pursue and why he or she is interested in the project. Students are not allowed to pursue something they already know how to do; they must choose something that will be a new learning experience for them.

### The Parts

**The Research Paper.** The research paper must include seven basic components before it is evaluated according to the criteria outlined on the "Senior Research Paper: Evaluation Sheet" (see Part 3 of this paper). The paper is first read by a South Medford English teacher, who makes comments but does not assign a grade. The student then makes revisions, and the paper is submitted to a reader from the local college, who grades it. Students who receive a failing grade on the paper have the opportunity to revise and submit the paper as many times as possible up to a given deadline—usually a two-week period. However, every effort is made to permit students to revise the paper until a satisfactory level of achievement is accomplished, which may require flexibility for completion deadlines.

In the letter of intent, students must also indicate that they understand the meaning of plagiarism and that it would result in failure of their project and no graduation if they engage in plagiarism.

FIGURE 1
ESSENTIAL LEARNING SKILLS

| Essential Learning Skills | | Senior Project | | |
|---|---|---|---|---|
| The student will be able to | | Research Paper | Project | Presentation |
| 1.3 | Speak with standard pronunciation, appropriate volume, gestures and inflections. | | | X |
| 2.1 | Identify main ideas, supporting details and facts/ opinions presented in oral and visual formats. | X | | X |
| 3.1 | Comprehend implied meanings of oral and visual communications. | X | | X |
| 4.1 | Determine the significance and accuracy of information and ideas presented in oral and visual communication | X | X | X |
| 4.4 | Listen, read, view, and evaluate presentations of mass media. | X | | |
| 6.1 | Recognize, construct, and draw inferences concerning relationships among things and ideas. | X | X | X |
| 6.2 | Generate and test interpretations, explanations, predictions and hypotheses. | X | X | |
| 6.3 | Identify problems and approach their solution in an organized manner. | | X | |
| 6.4 | Make reasoned evaluations of statements in mass media. | X | | |
| 7.1 | Clarify purposes of assignment. | X | X | X |
| 7.2 | Use resources beyond the classroom. | X | X | X |
| 7.3 | Select and use appropriate study techniques. | X | X | X |

**The Project.** The requirements for the project phase are much more loosely defined. The projects vary widely in their complexity and in the level of skill required. For example, the students may engage in a variety of activities, such as producing a product (a robot, a piece of Native American beadwork), learning a new skill (skiing, a vision quest), or providing a service (tutoring a student, organizing a community event to provide food for the homeless). The South Medford staff has found it exceedingly difficult to create criteria that would allow product evaluation while allowing students to pursue areas of interest that are challenging but not beyond their capability to perform.

Currently, the requirements include at least 15 hours out-of-school time and student self-reflection about the learning process experienced during the planning, problem solving, and completion of the chosen activity. It is possible that the project phase could be a "failure," i.e., the product created doesn't work, the vision quest lasts two instead of the planned four days, the tutored student still doesn't pass. If the student is able to analyze the learning process and determine what he or she has gained that is of value from engaging in the project phase, the student "passes" that part of the Senior Project.

**The Presentation.** The student's presentation is made in front of a panel of five judges, which includes faculty from the school and members of the community. The panelists are chosen because they have some knowledge, interest, or expertise in the area represented by the student's research paper and product. This culminating activity is rated by all five judges on the Senior Project Presentation Evaluation Form. The part that is difficult to measure is the excitement, enthusiasm, and mutual support generated by this event. Students take it very seriously, prepare accordingly, and celebrate their successes.

### The Forms

The forms presented on the following pages are based on those used by South Medford High School, but they have been edited for more general application. The collected data are useful for assessment and evaluation of student performance on the desired outcomes as stated earlier. Keep in mind that any forms developed and used by schools and districts would need to reflect local criteria and standards. The following forms are presented here:

**Senior Project Consent Form.** This form confirms the parent's or guardian's knowledge about the

BEST COPY AVAILABLE

Senior Project and participation in the selection of the student's project.

**Senior Project Description.** The purpose of this form is to have the student describe the nature of the project and the steps undertaken to complete the project. The student is also asked to reflect on what he or she has learned during the process. It is hoped that this will provide a sense of self-evaluation and reflection on the outcomes of the project.

It would be appropriate to consider the project phase of the Senior Project as evaluation of student performance only when criteria have been established for judging the projects. After further development, it may be possible to approach this in several ways. For example, the student may negotiate with a teacher or community member the criteria for the project and a quality standard that would reflect acceptable performance. Alternatively, the staff may establish more generic criteria and standards for each of several different categories for student projects. There would be different criteria for products, performance, or service projects. There may also be a combination of the two where there are criteria by category and the student negotiates the standards depending on the project.

**Senior Research Paper Evaluation Sheet.** In order to obtain consistent scoring of the research paper, agreed-upon standards are needed for what constitutes an excellent, good, fair and poor paper in each of the three categories. Both students and readers should be familiar with papers that reflect expected quality.

In preparation for the paper, students receive instruction on the basic components of the paper. The grade on each section is computed by averaging the numerical values given to each indicator. Students and readers may refer to the Grading Standard following the form.

**Senior Project Presentation Evaluation Form.** Again, students and panelists need to know the standards for quality performances by which the presentations will be judged. Also, the form as presented would give the content scores three times more value than either of the other two sections, unless they are weighted according to the number of items. If the content is intended to be the most important factor, then that needs to be clearly understood by all participants. If value for all three sections is intended to be equal, then the scoring needs to be adjusted mathematically or by changing the number of items in each section.

**Senior Project Verification.** This form is used simply to have someone other than the student verify that the student did indeed engage in the activity as described and that it was a learning experience.

## SENIOR PROJECT CONSENT FORM

As a parent/guardian of a student at ______________________ High School, I am aware that my son/daughter must pass all three phases of the Senior Project to graduate: 1) the research paper, 2) the project, and 3) the presentation. For the project phase, my son/daughter has decided to:

I fully understand that this project selection is made independently of the staff and administrationof the high school. This project selection and approval is parent and student centered.

Briefly, describe what you son/daughter will be doing to meet the requirement for this project phase of the Senior Project.

Parent/Guardian Signature ______________________________________
Student Signature ____________________________________________
Date _________________

Far West EDGE 1988, not to be reproduced.

## SENIOR PROJECT DESCRIPTION

Name ______________________________ Date ______ Period ______

1. In at least 25 words describe your project.

2. Fill in the blanks:
   A. Estimated total hours spent on project ______
   B. List the major steps involved in your project and indicate the time involved in each task.

Step 1 ______________________________ Time __________
Step 2 ______________________________ Time __________
Step 3 ______________________________ Time __________
Step 4 ______________________________ Time __________
Step 5 ______________________________ Time __________

C. Date started __________
Date completed __________

D. List reference materials used for your project
a. ______________________________
b. ______________________________
c. ______________________________

E. List people consulted for input on your project and put a check by all people who could potentially help future seniors.

______________________________
______________________________
______________________________

Far West EDGE 1988, not to be reproduced

## SENIOR PROJECT DESCRIPTION—(continued)

3. Circle Yes—No—NA (Not Applicable)
   - A. Was it helpful to your project to interview community resources? Yes No NA
   - B. Have you asked for advice/comments about your project from people other than the community resource? Yes No NA
   - C. Did you utilize task analysis and time management principles? Yes No NA
   - D. Have you ever done this type of project before? Yes No NA
   - E. Would you recommend your project area for future seniors? Yes No NA

     If yes, explain:

     If no, explain:

4. Choose yes or no to answer the following questions and explain why in 25 words or less.
   - A. Would you feel threatened to show your project to an expert for evaluation?

     If yes, why?

     If no, why?

   - B. List three things you now know after completing this entire project:
     1.
     2.
     3.

Far West EDGE 1988, not to be reproduced.

## SENIOR PROJECT DESCRIPTION—(continued)

C. List all personal satisfactions or knowledge you gained from this project experience.

1.

2.

3.

D. What comments about your projects have you heard from students, parents, teachers, or community members? For example: "I never knew you knew so much about motors."

5. Respond to the following in the space provided.
   A. List problems you encountered.

      1.

      2.

      3.

      4.

   B. How do you feel about having your paper, project description, and evaluation form on file as a model for next year's students? Explain why or why not.

   C. What was the picture in your mind of your project before you started working on it? How does this picture square with the outcome you have experienced?

   D. What grade would you give yourself for your project? Justify the evaluation of your grade in at least 25 words. My grade: ______
   Justification:

   E. If given the opportunity, what would you do differently now that you speak from experience?

Far West EDGE 1988, not to be reproduced.

## SENIOR PROJECT RESEARCH PAPER
### Evaluation Sheet

Name ______________________________ Period ______

**Basic Required Components:**
In order for a paper to be graded it must contain the following basic components:

| YES | NO | |
|---|---|---|
| ___ | ___ | Title page |
| ___ | ___ | Thesis (intro/conclusion) |
| ___ | ___ | Correct documentation within the paper |
| ___ | ___ | Minimum 5 citings including primary source |
| ___ | ___ | Minimum 3 sources (not including encyclopedias) |
| ___ | ___ | Minimum paper length |
| ___ | ___ | Works cited page/documentation/works consulted |
| ___ | ___ | Types (correct format) |

**Use of sources (1 poor rating fails the paper)**

| | Excellent | Good | Fair | Poor |
|---|---|---|---|---|
| Document all researched information | ☐ | ☐ | ☐ | ☐ |
| Works cited/consulted properly within text | ☐ | ☐ | ☐ | ☐ |
| Enough outside info to count as a research paper, not a personal essay | ☐ | ☐ | ☐ | ☐ |
| Mixture of writer's own words with paraphrased and quoted sources | ☐ | ☐ | ☐ | ☐ |
| Sources and citings match | ☐ | ☐ | ☐ | ☐ |

Grade ______________________

Far West EDGE 1988, not to be reproduced.

## SENIOR PROJECT RESEARCH PAPER
Evaluation Sheet—(continued)

**Content/Organization (2 poor ratings fail the paper)**

| | Excellent | Good | Fair | Poor |
|---|---|---|---|---|
| Introduction/Thesis | ☐ | ☐ | ☐ | ☐ |
| Quality of content (degree of topic difficulty/depth) | ☐ | ☐ | ☐ | ☐ |
| Writing sophistication (varied sentences, mature vocabulary, creativity, smooth sentence flow) | ☐ | ☐ | ☐ | ☐ |
| Body organization (follows and supports thesis) | ☐ | ☐ | ☐ | ☐ |
| Transitions | ☐ | ☐ | ☐ | ☐ |
| Conclusion (thesis and summary) | ☐ | ☐ | ☐ | ☐ |

Grade ______________________________

**Mechanics/Format (2 poor ratings fail the paper)**

| | Excellent | Good | Fair | Poor |
|---|---|---|---|---|
| Title page/pagination | ☐ | ☐ | ☐ | ☐ |
| Margins | ☐ | ☐ | ☐ | ☐ |
| Capitalization | ☐ | ☐ | ☐ | ☐ |
| Spelling | ☐ | ☐ | ☐ | ☐ |
| Punctuation | ☐ | ☐ | ☐ | ☐ |
| Usage/grammar | ☐ | ☐ | ☐ | ☐ |
| Agreement (unclear reference, subject/verb, pronoun) | ☐ | ☐ | ☐ | ☐ |
| Correct sentences (no run-ons, fragments) | ☐ | ☐ | ☐ | ☐ |
| Written formal style (third person) | ☐ | ☐ | ☐ | ☐ |

Grade ______________________________

Far West EDGE 1988, not to be reproduced.

## SENIOR PROJECT RESEARCH PAPER
### Grading Standard

1. All papers must contain the basic components BEFORE they will be graded.

2. To pass, each paper must not exceed the minimum number of "poors." The grade will be determined by the following standard:

   Excellent = 4
   Good = 3
   Fair = 2
   Poor = 1

Any paper receiving an "F" in any one of the three areas must be rewritten and corrected.

Use of Sources (1 poor rating fails the paper):

| | |
|---|---|
| 24—22 | A |
| 21—20 | B |
| 19—18 | C |
| 17—16 | D |

Content/Organization (2 poor ratings fail the paper):

| | |
|---|---|
| 24—22 | A |
| 21—20 | B |
| 19—18 | C |
| 17—16 | D |

Mechanics/Format (2 poor ratings fail paper):

| | |
|---|---|
| 36—34 | A |
| 33—31 | B |
| 30—29 | C |
| 28—27 | D |

*Getting something done is an accomplishment; getting something done right is an achievement.*

Far West EDGE 1988, not to be reproduced

## SENIOR PROJECT PRESENTATION
Evaluation Form

Student's name ______________________________
Paper: ____________________ Project: ____________________
Starting Time: ____________________ Finishing Time: ____________________

### AREA I—CONTENT

| | HIGH | | | | LOW |
|---|---|---|---|---|---|
| 1. *Introduction* (stated or implied purpose of presentation with an interesting attention getter) | 5 | 4 | 3 | 2 | 1 |
| 2. *Body of Speech* Main ideas and points supported by accurate and appropriate details | 8 | 6 | 5 | 3 | 1 |
| *Application of Learning* (listen for self-discovery cues such as "I learned," "I plan to use this knowledge" or "I never knew") | 5 | 4 | 3 | 2 | 1 |
| *Organization* (not choppy; logical flow of ideas) | 5 | 4 | 3 | 2 | 1 |
| 3. *Conclusion* | 5 | 4 | 3 | 2 | 1 |
| 4. *Balanced Content* (not too much on research paper or too much on project/ within 8-10 minute limit) | 5 | 4 | 3 | 2 | 1 |
| 5. *Language Usage* (transitions, appropriate word choice, etc.) | 5 | 4 | 3 | 2 | 1 |

TOTAL CONTENT ____________

### AREA II—DELIVERY

| | | | | | |
|---|---|---|---|---|---|
| 1. *Eye Contact* (not just reading notes, looks around) | 5 | 4 | 3 | 2 | 1 |
| 2. *Effective Speech Techniques* (poise, posture, rate/volume of voice, gestures) | 5 | 4 | 3 | 2 | 1 |
| 3. *Dress* | 5 | 4 | 3 | 2 | 1 |
| 4. *Audio and/or Visual Aids* | 5 | 4 | 3 | 2 | 1 |

TOTAL DELIVERY ____________

### AREA III—QUESTION AND ANSWER PERIOD

Judges questions (write at least one question)

| | | | | | |
|---|---|---|---|---|---|
| 1. *Impromptu skills* (how fluently, confidently and easily does student answer) | 5 | 4 | 3 | 2 | 1 |
| 2. *Quality of student response* (knowledge and information) | 5 | 4 | 3 | 2 | 1 |
| 3. *Student Project risk factor* | 5 | 4 | 3 | 2 | 1 |

TOTAL CONTENT ____________

**TOTAL SCORE AREAS I-III** ____________

Judges' Comments: Comments are appreciated by students. Include anything that could be of future use to the student, such as "you could market your product if you would see Mr. Jones," "you did a fine job explaining a difficult process," or "your portfolio was especially effective." (Use back of this form for your comments, please.)

Far West EDGE 1988, not to be reproduced.

## SENIOR PROJECT VERIFICATION

Name ______________________________

Project ______________________________

You have been chosen to verify this student's efforts on his/her Senior Project. Since most of the time spent on the project phase of the assignment has been out of class, verification of the student's efforts is necessary. Please answer the following questions to help us evaluate his/her performance. Please keep in mind that this student's research paper has already been evaluated. This form refers to the physical project.

1. Can you verify that he/she spent at least 15 hours creating this project?

   Yes No

   Comments:

2. Have you seen this project at different stages of completion, not just the final phase?

   Yes No

   Comments:

3. What problems specifically did this student encounter and overcome?

Signature ______________________
Phone ______________________
Relationship to student ______________________
Date ______________________

Far West EDGE 1988, not to be reproduced.

## Part 3: Examples of Data Displays

The data presentations that follow are simulations of data that could be prepared if the school chose to aggregate the individual student data to the school level. There are many ways the data could be analyzed, but for our purposes graphs will be created only to exemplify the type of data that would be appropriate for a school-level profile document. This will be followed by suggestions for a further breakdown of data that could be used for diagnostic purposes if it were decided that curriculum refinement is needed to better prepare the students for their Senior Project.

### Example 1: Simulation Data—South Medford High School

**Senior Projects**
**Non-Completion Rates**

**Prologue:** In order to graduate from South Medford High School, a student must satisfactorily complete a Senior Project which consists of a research paper, a project, and a presentation.

**Data Display:**

**Number of Students**
**Not Completing a Senior Project**
**1987-1991**

| Year | Senior Class Enrollment | Non-Completion of Senior Project | Non-completion by Category | | |
|---|---|---|---|---|---|
| | | | Research Paper | Project | Presentation |
| 1991 | 306 | 2 | 2 | 1 | 3 |
| 1990 | 310 | 2 | 2 | 1 | 2 |
| 1989 | 309 | 3 | 2 | 2 | 3 |
| 1988 | 315 | 5 | 4 | 3 | 5 |
| 1987 | 319 | 8 | 7 | 5 | 8 |
| | | 23 | 17 | 12 | 23 |

*Note: Students who do not competently complete either the research paper or the project do not make a presentation at the Senior Boards and do not graduate.

**Narrative:**

1. Of the 23 who did not graduate from 1987-1991, 17 did not competently complete the research paper.
2. Twelve students out of 23 did not competently complete the project.
3. Over the five years, there has been a decline from 8 students in 1987 to 2 in 1991 who fail to graduate because of the Senior Projects.

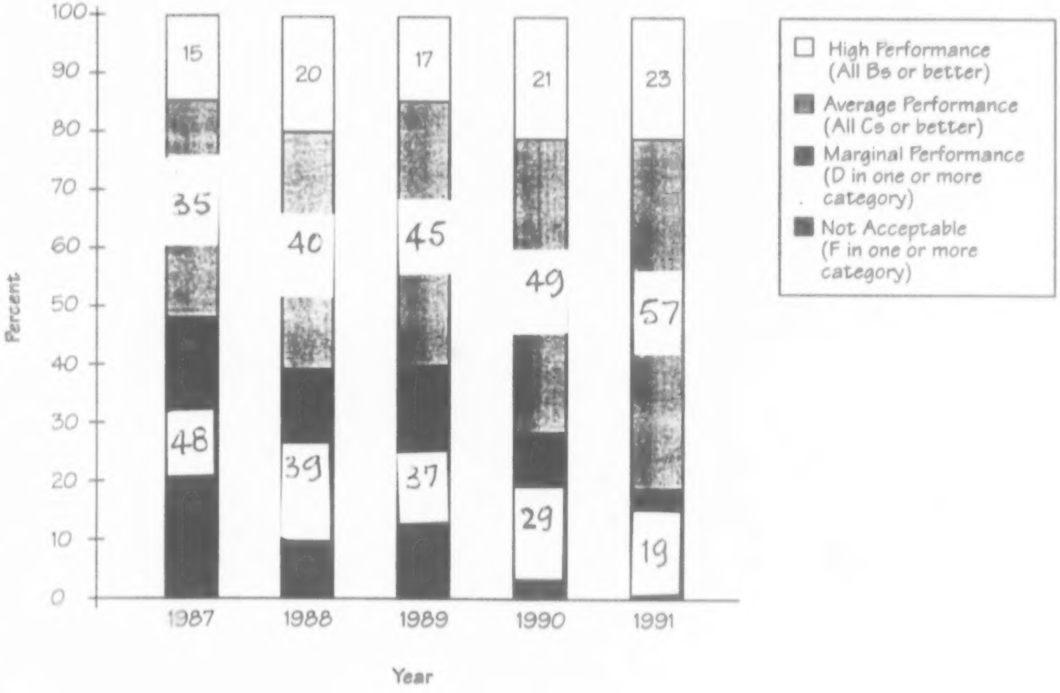
## Example 2: Simulation Data—South Medford High School

### Senior Projects
### Final Draft—Research Paper

**Prologue:** In order for the Senior Project research paper to be graded, seven basic components must be completed. (See Senior Research Paper: Evaluation Sheet). The paper is then given a letter grade based on a total score in each of the three scoring categories: I. Use of Resources, II. Content/Organization, and III. Mechanics/Format. Rewrites are permitted up to a given deadline.

**Data Display:**

**Percent of Students by Performance Category**
**Based on Grades**
**in All Three Scoring Categories**



**Narratives:**

1. The number of students receiving As and Bs in all Scoring categories has remained quite stable.
2. In five years, the number of students receiving a marginal performance has decreased from 48% to 19%.
3. Students receiving all Cs or better on all scoring categories has increased from 50% to 80% from 1987-1991.

BEST COPY AVAILABLE

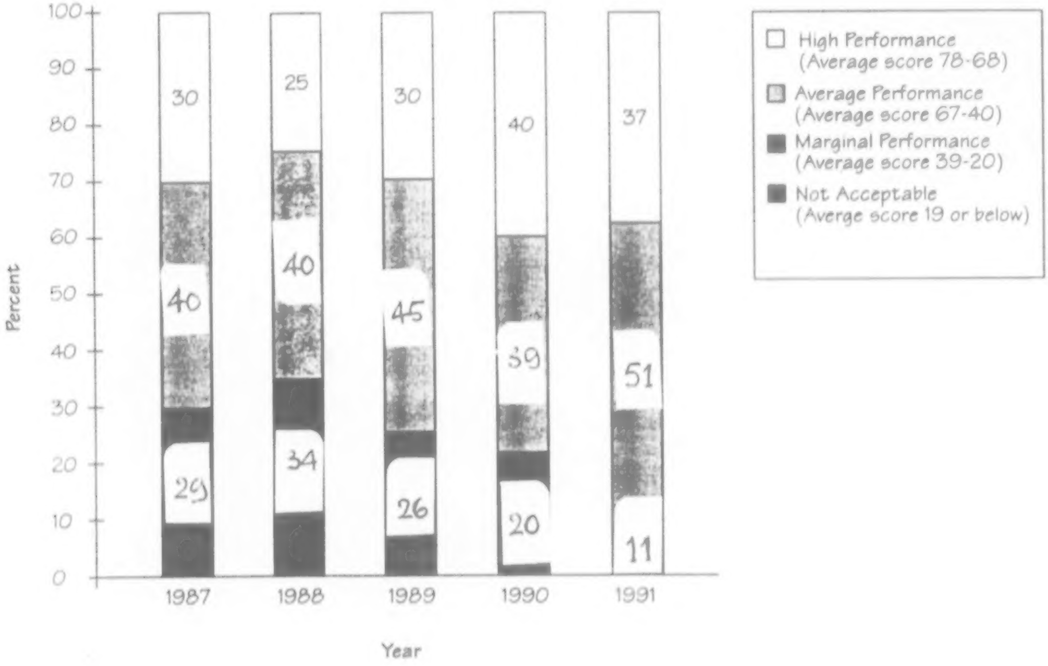
## Example 3: Simulation Data—South Medford High School

### Senior Projects
### Senior Boards Presentations

**Prologue:** After satisfactorily completing the research paper and the project, students participate in the presentation phase known as Senior Boards. A panel of 5 judges rate the students on a scale of 1 (low) to 5 (high) on 15 criteria in three areas: Area I—Content, Area II—Delivery, Area III—Question and Answer Period. The student receives a total score ranging form 78 to 15.

**Data Display:**

Percent of Students by Performance Category
Based on Average Score
of All Judges



**Narratives:**

1. Students scoring in the High Performance category range is increasing slightly over time.
2. Students scoring in the Marginal Performance category has decreased from a high 34% in 1988 to a low of 11% in 1991.
3. The highest percent of students scoring average or above was 88% in 1991.

BEST COPY AVAILABLE

## Example 4: Projects

As mentioned in Part 2, gathering and reporting data about the students' projects is much more difficult and subjective than gathering and reporting data about the research paper and presentation. Other than a required minimum of 15 hours for completing the project, there is no other set of criteria for judging the projects. It would be possible to determine the average number of hours students spend on their projects, but one cannot make any assumptions about quality based on time spent. Time could be a reflection of the difficulty of the project of it could also reflect a project gone awry due to lack of planning.

The purpose for doing the project is to assess the students' problem solving abilities, creativity, responsibility, and their ability to gain insight into their own learning process. The only evidence of this is through the student's self-reflection on the Project Description form. In addition, a parent or guardian is asked to complete a Project Verification form.

Parts of the Project Description could lend itself to analysis. For example, on Question 3 students are asked to answer Yes or No or NA to five questions that deal with their use of outside resources, time management, and whether or not the project was a stretch for them. Percentages of Yes or No answers could be tabulated.

Question 4-A, which asks students whether or not they feel threatened to show their project to an expert could be summarized into percentage of yes and no answers. It would be possible to analyze the comments to get a sense of their self-confidence about the projects. However, this is only one small piece of evidence and may reflect other student personality traits that cannot be measured here. The only possibility would be to look for trends to match against the general teacher perceptions. Interpretation of this information needs to be done with caution as further verification is needed.

Question 4-B asks students to list three things they know after completing the project. This information could be analyzed and categorized to identify whether what they learned had to do with knowledge and skills, application of knowledge, or the learning process and themselves as learners. The parents Project Verification form also asks parents to specify problems the student met and overcame. This information could also be categorized in a manner similar to those just mentioned. Question 5-D asks students to grade themselves on their projects and this information could be tabulated to see how many would assign themselves As, Bs, etc., and to see what criteria they cite since this has not been clearly defined for the projects.

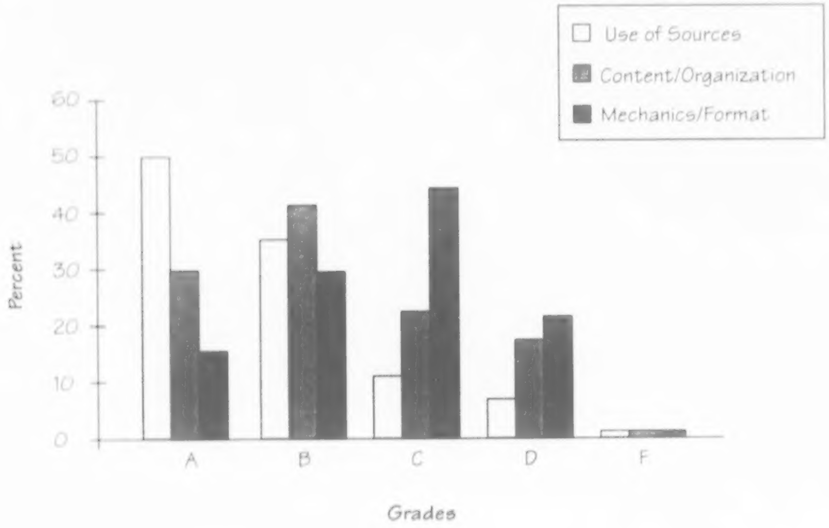
Basically, this type of analysis would be valuable to the staff as they document the data and see patterns emerge over time. A scheme could be developed for categorizing the types of projects and criteria could be developed for each categorial type. After a while it would be possible to see where the students are doing well and where they need assistance in the project phase of the Senior Projects. This kind of analysis is difficult at first when trying to identify the descriptors that can be used to categorize the information. Useful qualitative data may be obtained after careful study of this information.

## Example 5: Diagnostic Data

Diagnosis of curriculum and equity issues could be obtained through disaggregation of data such as the following:

1. Data on the research paper and the presentations disaggregated by sex, ethnicity, or socioeconomic status.
2. Data on the research paper disaggregated by the three categories: I. Use of Sources, II. Content/Organization, and III. Mechanics/Format

**Percent of Students**
**Receiving Letter Grades**
**by Category**



3. Data on the Senior Boards presentations disaggregated by the categories: Area I—Content, Area II—Delivery, Area III—Question and Answer Period.
4. Data about the attitudes on the effectiveness of the Senior projects collected by survey of students, staff, and community.

## Part 4: Issues and Concerns

If districts or schools are considering using an alternative assessment strategy as exemplified by the South Medford High School Senior Project, careful planning is required. Even though it seems to be the answer to a difficult curriculum alignment problem, this process may raise much anxiety for students, staff and parents alike. Before implementing such a dramatic change, there must be a lot of open discussion with all concerned parties. The Oregon Department of Education has identified many of the issues that need to be addressed (1988, pp. 23-29):

### Management and Recordkeeping Issues

1. **Reporting and recordkeeping:** The student proposal becomes the contract and is signed by the student, teacher, and parent or volunteer mentor.

*a. Individual student*

- Individual progress vs. group progress (if group projects are allowed, individual responsibilities would need to be identified)
- Student maintains and is responsible for record of own progress

*b. Classroom*

- Teacher monitors the projects for a single group of students
- Student portfolios are kept up-to-date by students and are reviewed regularly by the teacher

*c. Parent*

- Parents sign proposal/contract and sign at agreed checkpoints during each phase
- Parental involvement limits are clearly defined and parent/student roles outlined

*d. Building*

- Scheduling events (presentations) to accommodate all students, grades, project areas (i.e., cultural fair, science fair, writing festival)
- Maintaining permanent files (Phase III data has equal status with standardized test scores)
- Requesting and collecting teacher reports (teacher feedback and analysis of group and individual performance should reflect both student and program strengths and weaknesses)
- Aggregating data for district to match the Essential Learning Skills and the Common Curriculum Goals

*e. District*

- Aggregating data for building and state to provide buildings with a perspective for program improvements and staff development and to provide the state with a perspective for general improvements
- Follow-up building requests for assistance with staff and program development

2. **Manageability and practicality for classroom teachers:** an assessment of K-12 program requirements to meet these criteria:

- Simple (uncomplicated, clear, concise, and to the point)
- General (allows for student diversity of skills and knowledge without compromising important skill and knowledge and standards)
- Meaningful (important to the student and to teachers)
- Feasible (plans, procedures, time lines, and within scope of resources and facilities)
- Manageable (practical in that it does not add to program and personnel demands, even though it may change current practices)

3. **Responsibility and accountability for all participants:** similar, but separate and specific instructions will be needed for:

- Teachers—how to manage
- Students—how to succeed
- Parents—how to help
- Evaluators—what criteria to use and how to score

4. **Clear, comprehensive model:**

- Supported by examples from a variety of grade levels
- Supported by research reports and findings
- Supported by questions and answers on management issues from participating teachers
- Supported by change models like Onward to Excellence and Concerns Based Adoption Model
- Supported by alignment strategies with current practices like writing assessment, science fair, etc.

5. **Evaluators:**

- Students—peer evaluation and self-evaluation
- Teachers—progress checking
- Parents/mentors—progress checking
- Panels of judges—criteria minimum standards

6. **Use of classroom assessment:**

- Revising the topic and focus
- Renegotiating plan for completing the project
- Increasing the amount of assistance given to a student
- Changing the kind of assistance given to a student
- Individualizing instruction for some part of the project
- Regrouping students
- Reteaching to large or small groups
- Consistent accountability

**7. Uses of classroom results at building or district level:**

- Aggregate trait data from writing, visual, and oral presentations
- Cumulative files—maintaining portfolios of performance samples
- Teacher reports and observations about curriculum, instruction, assessment, and management choices
- Group agreements
- Curriculum (program) changes
- Staff development (instruction) needs

**8. District and state uses of assessment:**

- Formalization of procedures and standards
- Aggregating data for personal and public accountability
- Collecting models and providing training

**9. Assessment issues:**

Judy Arter, of the Northwest Regional Educational Laboratory Evaluation and Assessment Program, recommends that when planning assessment strategies, questions such as the following should be asked:

- Are there clear/appropriate criteria for judging student performance?
- Is there rater consistency and lack of bias?
- Does the sample represent what a student is capable of doing?
- Is the task biased, i.e., are some cultural groups at a disadvantage?
- Does the task invite real student performance?

## Summary

Schools and school districts across the nation are asking the question "What do we want high school students to learn?" Even more revealing than stated learner outcomes might be to look at what is expected from students when their time is up and how it is measured. In *Horace*, the Coalition of Essential Schools newsletter, Kathleen Cushman says that what students know and what they can do is in many ways a reflection of what their schools have expected them to master after a course or even a high school career is completed. Proving themselves through a series of mechanized hoops—machine-scanned textbook tests, achievement and competency tests—may often be the only evidence provided of this mastery (1990, p. 1).

In increasing numbers of schools, however, there is a belief that:

> A true test asks students to show what they know and can do, not to spout unrelated facts they have memorized the night before. Once we start measuring performance this way, change is swift to follow—in what we teach, in who we teach, and in our assumptions about why kids are in school at all (Cushman 1990, p. 1).

The South Medford High School Senior Project is a very good example of this emerging trend toward a more meaningfully aligned curriculum, where what we measure more closely reflects what we think is important for students to know and teachers to teach.

### REFERENCES

Cushman, Kathleen, Ed. (March 1990). *Horace* vol. 6, 3. Providence, R.I.: Coalition of Essential Schools.

Oregon Department of Education. (September 1988). *English Language Arts, Report 8,*

Integrated Assessment Model: A Project- Based Approach.

# X

## *Resources*

# X-1

# Glossary of Assessment Terms

COMPILED BY

JUDITH A. ARTER AND ROBERT E. BLUM
NORTHWEST REGIONAL EDUCATIONAL LABORATORY

**Alignment**

The process of strengthening the linkage among objectives, instructional materials, instructional methods, and assessments.

**Alternative Assessment**

Any type of assessment in which students create a response to a question (as opposed to assessments in which students choose a response from a given list, such as multiple-choice, true/false, or matching.) Alternative assessments can include short-answer questions, essays, performance assessments, oral presentations, demonstrations, exhibitions, and portfolios.

*See also* Authentic Assessment.

**Analytical Trait Scoring**

A scoring procedure in which performances are evaluated for selected dimensions or traits, with each trait receiving a separate score. For example, a piece of writing may be evaluated according to organization, use of details, attention to audience, and language usage/mechanics. Trait scores may be weighted and/or totaled.

*See also* Primary Trait Scoring.

**Anchor**

Example of performance that serves as a standard against which other performances may be judged; there are usually anchors showing examples of performance at different points on a scoring rubric. In math problem solving, for example, anchor papers are selected from actual student work exemplifying the quality of performance levels 1, 2, 3, and so forth. If used with analytical scoring, anchor papers or benchmark performances may be identified for each trait being assessed.

*See also* Benchmark, Rubric.

**Assessment**

The process of quantifying, describing, or gathering information about performance.

*See also* Evaluation.

**Authentic Assessment**

Assessment tasks that elicit demonstrations of knowledge and skills in ways that resemble "real life" as closely as possible. An authentic assessment also engages students in the activity and reflects best instructional activities. Thus, teaching to the authentic assessment is desirable.

**Belief or Value Statement**

Direct statement of commonly held beliefs about students, staff, schools, and other important aspects in the schooling process. Such statements answer the question "What do we believe about people, conditions, and control in the schooling process?" Two examples of belief statements follow:

- Every student can enjoy success as a learner and master the curriculum of the school at a high level of competency.
- We control the critical elements for learning success, including the instructional process, classroom and school environments.

**Benchmark**

An interpretation of a performance standard according to age, grade, or developmental levels. It could refer to samples of student work that illustrate excellent or adequate performance, or to reachable targets at various grade levels.

*See also* Anchor, Outcome, Performance Standard.

**Context (of a performance assessment)**

The circumstances in which a performance is embedded. For example, problem solving can be assessed in the context of a specific subject (e.g., mathematics) or in the context of a real-life laboratory problem requiring the use of mathematical, scientific, and communication skills; science process skills can be assessed in the context of a large-scale, high-stakes assessment or in the context of a classroom test.

**Criteria**

Guidelines, rules, or principles by which student responses, products, or performances are judged. Criteria may be holistic, analytic trait, general, or specific. The elements we value in student performance become rubrics or scoring guides when written down and used in assessment. Rubrics embody and express our criteria. For example, our criteria for good reading might include elements such as fluency, flexibility, making connections within and between tests, making connections with personal experience, and risk taking. These criteria would be expressed in rubrics, and the rubrics used to assess student performance in reading. Anchor papers or benchmark performances may be used to identify each level of competency in the rubric or scoring guide.

*See also* Analytical Trait Scoring, Benchmark, Criterion-Referenced Assessment, Evaluation, Holistic Scoring, Rubric

**Criterion-Referenced Assessment**

Comparing a student's performance according to a description of the desired performance. For example: She typed 55 words per minute without errors. This type of assessment is frequently contrasted with norm-referenced assessment.

*See also* Criteria, Norm-Referenced Assessment

**Dispositions**

Affective outcomes such as flexibility, perseverance, self-confidence, and positive attitude. Some new assessments attempt to measure these outcomes.

**District Priorities**

Statements that direct staff development and support efforts to enhance the potential of schools to achieve the mission of the district. District priorities address the question "What are the most important areas of staff activity that will lead toward accomplishment of the mission?" Here are some examples:

- During the next three years, the math curriculum, including objectives, resources, strategies, and assessments, will be revised to reflect the latest research in mathematics, mathematics teaching, and general effective teaching.
- Within the next three years, every school will initiate and/or maintain a school improvement process that has a primary focus on improving student performance, involves the full staff in setting goals for improvement, uses effective schooling research in developing improvement plans, involves the full staff in inservice activities as necessary to achieve goals, monitors and adjusts as necessary during the improvement process, and reports on improvement efforts annually.
- Within the next two years, procedures and instruments for assessing students' attitudes about the school as a place to learn and themselves as learners will be selected and/or developed.

**Evaluation**

Judgment regarding the quality, value, or worth of a response, product, or performance based on established criteria. Evaluations are usually based on multiple sources of information. The terms "evaluation" and "assessment" are often used interchangeably.

*See also* Assessment

**Expanded Definition (of individuals demonstrating learning success)**

Narrative descriptions of individuals who are demonstrating the characteristics and qualities intended by student learning goals (outcomes). The descriptions add specific meaning to the goals and are one step in the process of establishing shared meaning of student learning goals by individuals throughout the school district and the community.

Clarifying statements are the link between student learning goals (outcomes) and indicators. Examples of goals and expanded definitions follow:

- **Goals**: read and communicate effectively; think critically and solve problems; recognize and celebrate cultural diversity.
- **Expanded Definition**: When adolescents graduate from our school system, they will have a significant intellectual maturity. They will be able to analyze problems and issues, examine the component parts, and reintegrate them into either a solution or a new way of stating the problem or issue. In developing thinking skills, the youth will master self-expression and be able to "hear" others' expressions through diverse media. These skills of self-expression and hearing include persuasive and coherent writing, articulate verbal expression, and familiarity with symbols and basic vocabularies of the arts, mathematics, and the sciences. Moreover, students will be able to appreciate and absorb the perspectives of cultures (and languages) different from their own [Carnegie Corporation of New York, *Turning Points: Preparing American Youth for the 21st Century* (New York: Author, June 1989), p. 15].

### Extraneous Interference

Factors that might cause us to mismeasure students—for example, excessive reading on a math test, or role playing on a science assessment.

### Formative Assessment

Ongoing diagnostic assessment providing information to guide instruction and improve student performance.

*See also* Summative Assessment

### Generalizability

The extent to which the performances sampled by a set of assessment items or tasks are representative of the broader domain being assessed. For example, can we generalize about the student's problem-solving ability from the student's performance on a specific set of ten problem-solving tasks?

*See also* Sampling

### Holistic Scoring

Assigning one overall score to a performance.

*See also* Analytical Trait Scoring

### Improvement Target

Statement of the amount of improvement in a goal (outcome) area expected within a relatively short period of time. Targets are incremental improvements in reaching school performance standards. Improvement targets answer the question "How much improvement in our goal (outcome) area(s) can we achieve within the next year?" Examples follow:

- By the end of the 1988–89 school year, the percentage of students in the bottom quartile on the standardized test in reading will decrease from 32 to 25 percent.
- By the end of the 1988–89 school year, the average number of disciplinary referrals per week for serious offenses (fighting, stealing, drug use, etc.) will be reduced from 72 (the 1988–89 weekly average) to 50.
- By the end of the 1989–90 school year, the composite average NCE score on the standardized test will increase from 35 to 45.

### Indicator

A measure yielding statistics that describe performance related to student learning goals (outcomes) and other aspects of educational systems. An indicator must have a common, agreed upon, consistent definition and a reference point or standard against which performance can be judged. An indicator must meet technical standards of quality, such as measuring what is intended to be measured (validity) and measuring consistently (reliability) [J. Oakes, *Educational Indicators: A Guide for Policymakers*. CPRE Occasional Paper Series (New Jersey: Center for Policy Research in Education, Rutgers University, October 1986), pp. 1-2].

*See also* Outcome

### Interdisciplinary or Integrated Assessment

Tasks that assess students' abilities to apply concepts, principles, and processes from two or more subject disciplines to a central question, theme, issue, or problem.

### Large-Scale Assessment

Assessment for purposes and audiences outside the classroom. Large-scale assessments are usually standardized to some extent and given to large numbers of students at the same time. Purposes for large-scale assessment include accountability, program/curriculum evaluation, certification of student competency (e.g., high school graduation), and selection of students for special programs.

### Learning Goals

Statements describing the general areas in which students will learn and achieve. They typically reflect what students are expected to have learned by the time they leave high school. These statements answer the question "What should students have learned by the time they leave high school?" Examples follow:

- Read and communicate effectively
- Think critically and solve problems
- Develop positive self-concept, respect for others, and healthy patterns of behavior

- Work effectively in groups as well as individually
- Show appreciation for the arts and creativity
- Demonstrate civic, global, and environmental responsibility
- Recognize and celebrate multicultural diversity
- Exhibit technological literacy
- Have a well-developed knowledge base that enhances understanding and decision making
- Have positive problem-solving and thinking skills.

**Mission Statement**

A description of the basic purpose of the organization. It answers the question "What is the business of the organization?" For example, "The mission of the *(name)* School District is to assure that all students served by the district learn the skills, knowledge, behaviors, and attitudes necessary for productive living in a democratic society."

**Norm-Referenced Assessment**

Comparing a student's performance or test result to how other similar students did. For example: She typed better than 80 percent of her classmates. This type of assessment is frequently contrasted with criterion-referenced assessment.

*See also* Criterion-Referenced Assessment.

**On-Demand Assessment**

Assessment that takes place at a predetermined time and place. State tests, SATs, and most final exams are examples of on-demand assessment.

**Open-Ended Task**

A performance task having no single correct response. For example: "Below you see a bar graph without any labels. What might this be a graph of?" or "Here is an aquarium. You have $25 to buy fish, plants, rocks, etc. Use the attached information to plan what you will put in your aquarium. Justify or explain your choices." Open-ended tasks are usually used when the goal is to assess reasoning, critical thinking, and group process skills rather than specific knowledge.

**Open-Response Task**

A performance task having a single correct answer. For example: "There are four pieces of wood, each measuring seven feet. If you used them as a fence around your square yard, how large an area would you create?" or "Using the following order form and catalog, order the 'Southwest Hat' in size 9, 'work boots,' and 'Jersey windbreaker.' "

**Outcome**

A goal statement specifying desired knowledge, skills or processes, and attitudes to be developed as a result of educational experiences.

*See also* Indicator, Learning Goals

**Performance Assessment**

Direct observation of student performance or student work and professional judgment of the quality of that performance. Good-quality performance assessment has preestablished performance criteria.

*See also* Criteria

**Performance Criteria**

*See* Criteria

**Performance Standard**

An established level of achievement, quality of performance, or degree of proficiency.

*See also* Anchor

**Portfolio**

A purposeful, integrated collection of student work showing effort, progress, or achievement in one or more areas. Usefulness for instruction and assessment is enhanced when students select the items for their portfolios, self-reflection is encouraged, and criteria for success are clear.

**Primary Trait Scoring**

A scoring procedure by which products or performances are evaluated by limiting attention to a single criterion or a few selected criteria. These criteria are typically based on the trait or traits that are most essential to a good performance. For example, if a student is asked to write to the Department of Energy to urge the opening or closing of a nuclear power plant, the primary traits might be the ability to communicate persuasively and the correct application of scientific knowledge to back up one's position. Scorers would attend only to these two traits.

*See also* Analytical Trait Scoring, Holistic Scoring

**Proficiency**

Having or demonstrating a high degree of knowledge or skill in a particular area.

**Prompt**

An assignment or directions asking the student(s) to undertake a task or series of tasks. Complete prompts present the context of the situation, the problem(s) to be solved, and criteria or standards by which responses will be evaluated.

**Reliability**

The degree to which the results of an assessment are dependable and yield consistent results. Reliability is an indication of the consistency of scores across evaluators, over time, or across different tasks that measure the same thing. An assessment is considered reliable when the same answers receive the same score no matter when the assessment occurs or who does the scoring. Technically, reliability is a statistical term that defines the extent to which errors of measurement are absent from an assessment instrument.

*See also* Generalizeability, Validity

**Rubric**

An established set of criteria for scoring or rating students' performance on tasks. Good rubrics consist of a fixed measurement scale (e.g., 4-point), a description of the characteristics of products or performances being measured for each score point, and sample responses (anchors) that illustrate the various levels of performance.

*See also* Anchor

**Sampling**

The selection of an array of performances to be assessed that is wide enough in range and large enough in number to: (a) cover the scope of the performances addressed by the assessment system in terms of representativeness and comprehensiveness, and (b) permit valid inferences about performances to be made. For example, the physical condition of a firefighter might be assessed by asking the firefighter to perform a series of activities while in full gear rather than looking at every possible physical task that might have to be performed as part of the job.

**Scale**

The range of scores possible on a test or individual performance task. Performance assessment items are typically scored on a four- to six-point scale, compared with a scale of two (right/wrong) on multiple-choice measures.

**School Improvement Goals**

Statements developed at each school to describe improvements in student performance to be achieved by the school. School improvement goals answer the question "What improvements in student performance will we achieve in this school over the next year or years?" Examples follow:

- Over the next three years, the percentage of students above the 50th percentile in reading on the Stanford Achievement Test will increase from 30 percent to 50 percent.
- By the end of the 1988–89 school year, average daily attendance will increase from 82 percent to 90 percent and the number of chronic absentees (students being absent more than 20 days per semester) will be reduced from 10 to 7 percent of the student body.
- Over the next three years, the percentage of students in the superior range on a holistically scored writing sample will increase from 2 to 10 percent, and the percentage of students in the unacceptable range will decrease from 40 to 20 percent.

**School Performance Standards**

Statements of acceptable levels of attainment by the students in the school. They are specific at the school level and represent a broad range of areas of performance. School performance standards answer the question "How well do we expect students to be performing in our school?" Examples follow:

- The average NCE score on standardized math tests will be 85 percent in the *(name)* school.
- In the *(name)* school, at least 40 percent of the students will score in the outstanding range on a holistically scored writing sample.
- Daily attendance in the school will be maintained at 95 percent.

**Standardization**

A set of consistent procedures for constructing, administering, and scoring an assessment. The goal of standardization is to ensure that all students are assessed under uniform conditions so that interpretation of their performance is comparable and not influenced by differing conditions. Standardization is an important consideration if comparisons are to be made between scores of individuals or groups.

**Summative Assessment**

Culminating assessment for a unit, grade level, or course of study providing a status report on mastery or degree of proficiency according to identified learning goals (outcomes).

*See also* Formative Assessment.

**Task**

An activity, exercise, or problem given to the students to perform. The best tasks require multiple responses to a challenging question or problem and are aligned with one or more student learning goals (outcomes). For example, students write an essay on an assigned topic, design an experiment, or solve a math problem.

*See also* Assessment, Performance Assessment, Test.

**Test**

A set of questions or situations designed to permit an inference about what an examinee knows or can do in one or more goal (outcome) areas.

*See also* Assessment, Performance Assessment, Task.

**Validity**

Refers to whether or not an assessment measures what it was intended to measure and nothing more. For example, a valid assessment of mathematical problem solving would be to measure the students' ability to solve a problem, not the ability to read the problem. Technically, validity indicates the degree of accuracy of predictions or inferences based on an assessment measure.

*See also* Generalizability, Reliability.

### Vision Statement

A description of an ideal situation to be achieved. A vision statement should include descriptions of students who have learned well, and it may include descriptions of ideal staff, working conditions, facilities, and so on. Examples follow:

- Students leaving *(name)* High School after 13 years in the *(name)* School District will be self-assured individuals with the next steps in their careers planned and organized. They will have the language skills—reading, writing, listening, and speaking—to participate in personal, occupational and social activities, including the electoral process.
- Within five years, central office and school staff members will have a basic understanding of effective schooling research and be using the research to plan ways to improve student learning in the schools. In every school, the faculty, administration, and others will be learning collegially and applying their knowledge and skills in the day-to-day activities in the school and classrooms.

### Bibliography

Hart, D. (1993). *Authentic Assessment: A Handbook for Educators*. Reading, Mass.: Addison-Wesley.

McTighe, J. (1993). *Glossary of Assessment Terms*. Frederick, Md.: Maryland Assessment Consortium, c/o Frederick County Public Schools, 115 E. Church St., Frederick, MD 21701.

Regional Educational Laboratory Network Program. (1994). *Improving Science and Mathematics Education—A Toolkit for Professional Developers*. 2nd ed. Portland, Oreg.: Northwest Regional Educational Laboratory.

Stiggins, R.J. (March 1991). "Assessment Literacy." *Phi Delta Kappan* 72, 7: 534-539.

Wheeler, P., and G.D. Haertel. (1993). *Resource Handbook on Performance Assessment and Measurement: A Tool for Students, Practitioners, and Policymakers*. Berkeley, Calif.: The Owl Press.

# X-2

# Assessment Information Sources*

COMPILED BY

JUDITH A. ARTER
NORTHWEST REGIONAL EDUCATIONAL LABORATORY

VIDEO RESOURCES COMPILED BY
JAY MCTIGHE
MARYLAND ASSESSMENT CONSORTIUM

## Agencies

The agencies on this list are involved in alternative assessment at some level. If no specific description of an agency's activities is given, the agency is likely doing one or more of the following: developing content standards, developing alternative assessments, developing materials to assist others in the development of alternative assessments, or conducting training on alternative assessments. Including agencies and publishers on this list should not be taken as an endorsement of their services or products.

**American Association for the Advancement of Science (AAAS)**, 1333 H St. N.W., Washington, DC 20005. Phone: (202) 326-6400. Has information on alternative assessment and curriculum standards in science.

**American Educational Research Association (AERA)**, 1230 Seventeenth St., N.W., Washington, DC 20036. Phone: (202) 223-9485. Has a subcommittee working on assessment standards.

**Appalachia Educational Laboratory**, 1031 Quarrier St., P.O. Box 1348, Charleston, WV 25325. Phone: (304) 347-0400, Helen Saunders. Provides training and materials on alternative assessment.

**Assessment Training Institute**, 215 S.W. Washington, Suite 201, Portland, OR 97204. Phone: (503) 228-3060, Rick Stiggins. Provides training, videos, and print material on classroom assessment.

**Association for Supervision and Curriculum Development (ASCD)**, 1250 N. Pitt St., Alexandria, VA 22314. Phone: 1-800-933-2723 or (703) 549-9110. World Wide Web: http://www.ascd.org/. Coordinates consortium efforts in assessment development and publishes print and video resources.

---

*Some listings based on information in Patricia Wheeler and Geneva Haertel, *Resource Handbook on Performance Assessment and Measurement* (Berkeley, Calif.: Owl Press, 1993). Mailing address: P.O. Box 89, Berkeley, CA 94701.

**Buros Institute of Mental Measurements**, The University of Nebraska—Lincoln, P.O. Box 880348, 135 Bancroft Hall, Lincoln, NE 68588. Phone: (402) 472-6203. Conducts and publishes reviews of tests.

**Center for Civic Education**, 5146 Douglas Fir Rd., Calabasas, CA 91302. Phone: (818) 591-9321. Coordinates the effort to develop curriculum standards in civics.

**Center for the Study of Reading**, 174 Children's Research Center, 51 Gerty Dr., Champaign, IL 61820. Phone: (217) 333-2552. Coordinates the development of curriculum standards in English.

**Center on Learning, Assessment and School Structure (CLASS)**, 648 The Great Road, Princeton, NJ 08540. Phone: (603) 252-1211, Grant Wiggins. Provides training and videos on assessment.

**Coalition of Essential Schools**, Brown University, Box 1969, Providence, RI 02912. Phone: (401) 863-3384, Joe McDonald. Coordinates cooperative efforts on alternative assessment; disseminates *Exhibitions Collection*, a sampler of performance assessments.

**Council of Chief State School Officers (CCSSO)**, 1 Massachusetts Ave., N.W., Suite 700, Washington, DC 20001. Phone: (202) 336-7045, Ed Roeber; (371) 328-0639, Linda Bond. Conducts yearly surveys of state assessment activities.

**Council on Educational Development and Research (CEDaR)** 2000 L St., N.W., Suite 601, Washington, DC 20036. Phone: (202) 223-1593. Coordinates the nation's Regional Laboratories and Research Centers, several of which are involved in alternative assessment projects.

**Educational Testing Service (ETS)**, Rosedale Rd., Princeton, NJ 08541. Phone: (609) 921-9000. Provides training and materials on assessment; develops assessments under contract.

**ERIC Clearinghouse on Assessment and Evaluation**, Catholic University of America, 209 O'Boyle Hall, Washington, DC 20064. Phone: (202) 319-5120. Houses library of articles and information on assessment. The collection is available on-line (http://www.cua.edu/www/eric_ae/).

**Far West Laboratory for Educational Research and Development (FWL)**, 730 Harrison St., San Francisco, CA 94107. Phone: (415) 565-3000. Provides training and materials on assessment; develops assessments under contract.

**Laboratory Network Program (LNP)**. The 10 regional educational laboratories have collaborated to develop a science and math alternative assessment database and training materials. The 10 labs (see individual agency listings for contact information) are: Appalachia Educational Laboratory, Far West Laboratory, Mid-continent Regional Educational Laboratory, North Central Regional Educational Laboratory, Northwest Regional Educational Laboratory, Pacific Regional Educational Laboratory, Northeast and Islands Regional Educational Laboratory, Research for Better Schools, Southeast Educational Development Laboratory, and Southeastern Regional Vision for Education. The database is also available on Internet (http://www.nwrel.org). Some Labs have lending libraries.

**Lawrence Hall of Science**, University of California, Berkeley, CA 94720. Phone: (510) 642-4193. Publishes instructional materials and assessment instruments in science.

**Mid-continent Regional Educational Laboratory (McREL)**, 2550 S. Parker Rd, Suite 500, Aurora, CO 80014. Phone: (303) 337-0990.

**Music Educators National Conference**, 1902 Association Dr., Reston, VA 22091. Phone: (703) 860-4000. Coordinates development of curriculum standards in the arts.

**National Assessment of Educational Progress, (NAEP)**, Educational Testing Service, Princeton, NJ 08541. Phone: (609) 921-9000. Conducts national assessments of student achievement in reading, writing, mathematics, science, fine arts, and social studies. Some components are performance-based.

**National Assessment Governing Board (NAGB)**, 800 N. Capitol St., N.W., Suite 825, Washington, DC 20002. Phone: (202) 357-6938. Is the governing body for the *National Assessment of Educational Progress*.

**National Association for Sport and Physical Education**, 1900 Association Dr., Reston, VA 22091. Phone: (703) 476-3410. Coordinates development of curriculum standards in physical education.

**National Center for History in the Schools**, UCLA, 231 Moore Hall, 405 Hilgard Ave., Los Angeles, CA 90024. Phone: (310) 825-470. Coordinates development of curriculum standards in history.

**National Center for Research in Mathematical Sciences Education**, University of Wisconsin, 1025 W. Johnson St., Madison, WI 53706. Phone: (608) 263-4285, Don Chambers.

**National Center for Research on Evaluation, Standards and Student Testing (CRESST)**, UCLA Graduate School of Education, 145 Moore Hall, 405 Hilgard Ave., Los Angeles, CA 90024. Phone: (310) 206-1532, Ron Dietal. World Wide Web: http://www.cse.ucla.edu/CRESSTHome.html. Provides Internet database on alternative assessment in all subject areas; conducts training and research on alternative assessment.

**National Center for the Study of Writing**, University of California at Berkeley, 5513 Tolman Hall, Berkeley, CA 94720. Phone: (510) 643-7022. Houses the university-based National Writing Project (NWP) network consisting of over 160 training sites located at universities across the nation.

**National Council on Measurement in Education (NCME)**, 1230 Seventeenth St., N.W., Washington, DC 20036. Phone: (202) 223-9485. Has a working committee on assessment standards.

**National Council for the Social Studies**, 3501 Newark St., N.W., Washington, DC 20016. Coordinates development of curriculum standards in social studies.

**National Council of Teachers of English (NCTE)**, 111 Kenyon Road, Urbana, IL 61801. Phone: 1-800-369-NCTE. Regularly publishes information on alternative assessment in reading, writing, speaking, and listening.

**National Council of Teachers of Mathematics (NCTM)**, 1906 Association Dr., Reston, VA 22091. Phone: 1-800-235-7566. Regularly publishes information on alternative assessment and curriculum standards in mathematics.

**National Geography Education Standards Project**, 1145 17th St. N.W., Washington, DC 20036-4688. Phone: (202) 828-5653. Coordinates development of curriculum standards in geography.

**National Science Education Standards**, 2101 Constitution Ave., N.W., HA 486, Washington DC 20418. Phone: (202) 334-2000. Coordinates development of curriculum standards in science.

**New Standards Project**, University of Pittsburgh, Learning Research and Development Center, 3939 O'Hara St., Pittsburgh, PA 15213. Phone: (412) 624-8319. Coordinates cooperative efforts in alternative assessment.

**North Central Regional Educational Laboratory (NCREL)**, 1900 Spring Rd., Suite 300, Oak Brook, IL 60521. Phone: (708) 571-4700, Jeri Nowakowski. Publishes assessment information, including a database on state assessment activities and monographs on legal issues.

**Northeast and Islands Regional Educational Laboratory**, 300 Brickstone Square, Suite 900, Andover, MA. Phone: (800) 347-4200.

**Northwest Evaluation Association (NWEA)**, 5 Centerpointe Dr., Suite 100, Lake Oswego, OR 97035. Phone: (503) 624-1951, Allan Olson. Provides training in assessment, computer-adaptive testing.

**Northwest Regional Educational Laboratory (NWREL)**, Test Center, 101 S.W. Main St., Suite 500, Portland, OR 97204. Phone: (503) 275-9582. World Wide Web: http://www.nwrel.org. Houses lending library of alternative assessment instruments, information, and bibliographies; provides training on assessment.

**Office of Educational Research and Improvement (OERI)**, 555 New Jersey Ave., N.W., Washington, DC 20208-5644. Phone: (202) 219-1828. Funds many assessment projects.

**Pacific Regional Educational Laboratory**, 828 Fort Street Mall, Honolulu, HI 96813. Phone: (808) 533-6000.

**Portfolio Assessment Clearinghouse**, San Dieguito Union High School, 710 Encinitas Blvd., Encinitas, CA 92024. Phone: (619) 753-6491. Newsletter on portfolio projects.

**Research for Better Schools**, 444 N. 3rd St., Philadelphia, PA 19123. Phone: (215) 574-9300.

**Southeast Educational Development Laboratory**, 211 E. 7th St., Austin, TX 78701. Phone: (572) 476-6861.

**Southeastern Regional Vision for Education**, P.O. Box 5367, Greensboro, NC 27435. Phone: (800) 755-3277.

## State Departments of Education and School Districts

State departments of education and school districts are in the forefront of developing alternative assessments. Presence on this list does not necessarily imply endorsement of the assessments being developed.

**Alberta Education**, Box 43, 11160 Jasper Ave., Edmonton, Alberta, AB T5K 0L2, Canada. Phone: (403) 427-0010. Alberta has developed performance assessments for 6h and 9th graders in science and mathematics. Staff have also developed handbooks to assist teachers with classroom assessment.

**Connecticut Department of Education**, Division of Research, Evaluation and Assessment, Box 2219, Hartford, CT 06145. Phone: (203) 566-1684, Douglas Rindone. Connecticut has developed performance assessments in writing, science, math, reading, and social studies.

**Illinois State Board of Education**, 100 N. 1st St., Springfield, IL 62777. Phone: (217) 782-2221. Illinois was one of the first states to redesign its statewide reading assessment. They also have direct assessments in writing and math.

**Kansas State Board of Education**, 120 S.E. 10th Ave., Topeka, KS 66612. Phone: (913) 296-3201. Kansas has developed pilot open-ended math, science, reading, and social studies assessments for grades 4, 7, and 10.

**Kentucky Department of Education**, Office of Assessment and Accountability, 19th Floor, Capital Plaza Tower, 500 Mero St., Frankfort, KY 40601. Phone: (502) 564-4394, Edward Reidy. *The Ken-*

*tucky Instructional Results Information System* (KIRIS) is developing short-response items in reading, math, science, and social studies. They are also working on portfolios.

**Maryland Department of Education**, 200 W. Baltimore St., Baltimore, MD 21201. Phone: (410) 333-2000, Steven Ferrara. The *Maryland School Performance Assessment Program* (MSPAP) is a series of assessment tasks for grades 3, 5, and 8 which are integrated across reading, writing, science, math, and social studies. These are used for school accountability.

**Maryland Assessment Consortium**, c/o Frederick County Public Schools, 115 E. Church St., Frederick, MD 21701. Phone: (301) 694-1337, Jay McTighe. Consortium participants cooperatively develop performance assessments in science, math, reading, writing, and social studies; video and other training materials are available.

**Massachusetts Department of Education**, 1385 Hancock St., Quincy, MA 02169. Phone: (617) 770-7334. The *Massachusetts Educational Assessment Program* (MEAP) has developed open-ended assessments in reading, social studies, science, and math for grades 4, 8, and 12.

**Michigan Department of Education**, P.O. Box 420, Lansing, MI 48902. Phone: (577) 373-8393. Michigan designed an *Employability Skills Portfolio* to help students demonstrate that they have acquired the skills needed for employment.

**New York State Education Department**, Division of Educational Testing, Room 770 EBA, Albany, NY 12234. Phone: (518) 474-5902, Carolyn D. Byrne. New York has developed performance assessments in several areas, including writing and science. The state is developing a computerized Macintosh and IBM resource bank of performance assessments called the *New York Assessments Collection*.

**Oregon Department of Education**, Public Service Building, 255 Capitol St., N.E., Salem, OR 97310. Phone: (503) 378-8004, Barbara Wolf. Oregon has conducted writing assessments using a six-trait model since 1985 and is developing on-demand performance assessment, projects, and portfolios in grades 3, 5, 8, and 10 in science, reading, and math.

**Pittsburgh Public Schools (ARTS PROPEL)**, Division of Writing & Speaking, 341 S. Bellefield Ave., Pittsburgh, PA 15213. Phone: (412) 622-3500; fax: (412) 622-8578. ARTS PROPEL has developed portfolios in creative writing, music, and art.

**Texas Education Agency**, 1701 North Congress Ave., Austin, TX 78701. Phone: (512) 463-9536, Barbara Green. Texas has developed performance assessments for social studies and science in grades 4 and 8.

**Vermont Department of Education**, 120 State Street, Montpelier, VT 05620. Phone: (802) 828-3135, Richard P. Mills, Commissioner. Vermont is developing portfolios for grades 4 and 8 in writing and math problem solving. The purpose is program evaluation and improvement.

## Publishers

Many test publishers have performance assessment components available for their achievement test series. These range from short fill-in-the-blank questions to more in-depth, integrated tasks. Norms are frequently available. Presence on this list does not necessarily imply endorsement of products.

**American Guidance Service**, 41 Woodland Road, P.O. Box 99, Circle Pines, MN 55014. Phone: 1-800-328-2560.

**CTB/McGraw-Hill**, P.O. Box 150, Monterey, CA 93942. Phone: 1-800-538-5380. Publishes standardized multiple-choice tests (CTBS, CAT) and related performance assessments.

**Educational Testing Service (ETS)/National Assessment of Educational Progress (NAEP)**, Educational Testing Service, Princeton, NJ 08541. Phone: 1-800-223-0267. The NAEP is administered yearly to a national sample of students. Some assessments have performance components. ETS is also involved in a number of other assessment projects.

**IOX Associates**, 5301 Beethoven St., Suite 109, Los Angeles, CA 90066. Phone: (310) 822-3275. Publishes training videos on assessment.

**Psychological Corporation**, 555 Academic Court, San Antonio, TX 78204. Phone: 1-800-228-0752. Publishes standardized multiple-choice tests (SAT, MAT) and related performance assessments (GOALS—short answer; *Integrated Assessment System*—longer, more integrated tasks).

**Riverside Publishing Company**, 8420 Bryn Mawr Ave., Chicago, IL 60631. Phone: 1-800-323-9540. Publishes standardized multiple-choice tests (ITBS, TAP, ITED) and related performance assessments.

**Touchstone Applied Science Associates (TASA)**, Fields Lane, P.O. Box 382, Brewster, NY 10509. Phone: (914) 277-4900. Publishes *Degrees of Reading Power*, which uses cloze questions.

## Print Materials

### NCME Instructional Modules

Since 1987, the National Council on Measurement in Education (NCME) has provided a series of instructional units, called ITEMS, in its quarterly publication *Educational Measurement: Issues and Prac-*

*tice*. The list of ITEMS through fall 1993 appears below:

Arter, Judith A., and Vicki Spandel. (Spring 1992). "Using Portfolios of Student Work in Instruction and Assessment." 11, 1: 36-44.

Brennan, Robert L. (Winter 1992). "Generalizability Theory." 11, 4: 27-34.

Brookhart, Susan. (Spring 1993). "Assessing Student Achievement with Term Papers and Written Reports." 12, 1: 40-47.

Cook, Linda L., and Daniel R. Eignor. (Fall 1991). "IRT Equating Models." 10, 3: 37-45.

Frary, Robert R. (Summer 1988). "Formula Scoring of Multiple Choice Tests (Correction for Guessing)." 7, 2: 33-38.

Frisbie, David A. (Spring 1988). "Reliability of Scores from Teacher-Made Tests." 7, 1: 25-35.

Frisbie, David A., and Kristie K. Waltman. (Fall 1992). "Developing a Personal Grading Plan." 11, 3: 35-42.

Hambleton, Ronald K., and Russell W. Jones. (Fall 1993). "Comparison of Classical Test Theory and Item Response Theory." 12, 3: 38-47.

Harris, Debrah. (Spring 1989). "Comparison of 1-, 2-, and 3-Parameter IRT Models." 8, 1: 35-41.

Harvill, Leo M. (Summer 1991). "Standard Error of Measurement." 10, 2: 33-41.

Kolen, Michael J. (Winter 1988). "Traditional Equating Methodology." 7, 4: 29-36.

Oosterhof, Albert C. (Winter 1987). "Obtaining Intended Weights When Combining Student Scores." 6, 4: 29-37.

Osterlind, Steven. (Fall 1988). "Using CRTs in Program Curriculum Evaluation." 7, 3: 23-30.

Stiggins, Richard J. (Fall 1987). "Design and Development of Performance Assessments." 6, 3: 33-42.

Stiggins, Richard J. (Summer 1992). "High Quality Classroom Assessment: What Does It Really Mean?" 11, 2: 35-39.

Traub, Ross E., and Glenn L. Rowley. (Spring 1991). "Understanding Reliability." 10, 1: 37-45.

### Sources of Information on Tests

Conoley, J.C., and Impara, J.C., eds. (1994). *The Supplement to the Eleventh Mental Measurements Yearbook*. Lincoln: University of Nebraska Press.

Educational Testing Service. *The ETS Test Collection Catalog*. Phoenix, AZ: Oryx Press, 1989-92.

- Volume 1: Achievement Tests (2nd ed.), 1993
- Volume 2: Vocational Measures, 1995
- Volume 3: Tests for Special Populations, 1989
- Volume 4: Cognitive Aptitude Tests, 1990
- Volume 5: Attitude Tests, 1991
- Volume 6: Personality Tests, 1992

Keyser, Daniel J., and Richard C. Sweetland, eds. (1984-92). *Test Critiques (Volumes I-IX)*. Austin, Texas: PRO-ED, Inc.

Kramer, Jack J., and Jane Close Conoley, eds. (1992). *The Eleventh Mental Measurements Yearbook*. Lincoln: The University of Nebraska—Lincoln, Buros Institute of Mental Measurements.

Mitchell, James V., Jr. (1994) *Tests in Print*. 3rd ed. Lincoln: The University of Nebraska—Lincoln, Buros Institute for Mental Measurements.

Sweetland, Richard C., and Daniel J. Keyser, eds. (1991). *Tests: A Comprehensive Reference for Assessments in Psychology, Education, and Business*. 3rd ed. Austin, Texas: PRO-ED, Inc.

### Articles and Books

#### General Changes in Assessment, Instruction, and Goals for Students:

American Federation of Teachers, National Council on Measurement in Education, and National Education Association. (Winter 1990). "Standards for Teacher Competence in Educational Assessment of Students." *Educational Measurement: Issues and Practice* 9: 30-32. The standards represent a conceptual framework or scaffolding from which specific skills can be derived.

Bond, Linda A. (1993). *Executive Summary of Issues and Recommendations Regarding Implementation of High School Graduation Tests*. Oak Brook, Ill.: North Central Regional Educational Laboratory, Regional Policy Information Center. This paper focuses on the practical policy issues that a state should consider as it adopts a high school graduation test.

Educational Testing Service. (1994). *Policy Information Report—Testing in America's Schools*. Princeton, N.J.: ETS, Policy Information Center. Describes the types and prevalency of assessments being undertaken by states.

Gardner, Howard. (July 1993). "Educating for Understanding." *The American School Board Journal* 180, 7: 21-24. A good overview of the changing goals we have for students.

Herman, Joan L., Pamela R. Aschbacher, and Lynn Winters. (1992). *A Practical Guide to Alternative Assessment*. Alexandria, Va.: ASCD. A step-by-step approach to K-12 assessment development with examples of assessment tasks and important issues to consider.

Hymes, Donald. (1991). *The Changing Face of Testing and Assessment: Problems and Solutions. AASA, Critical Issues Report*. Arlington, Va.: American Association of School Administrators. A good overview of changes in assessment and the reasons for them.

McDaid, Janet L., and Donna G. Davis. (October 1991). "Using Assessment Results Wisely." *Thrust*

*for Educational Leadership* 21, 2: 34-36. Considerations when designing an assessment system, including the roles and responsibilities of various interested parties.

Murray-Ward, Mildred, and Annie W. Ward. (1993). *Assessment for All Students: Where Have We Been and Where Are We Going?* Thousand Oaks, Calif.: California Lutheran University. Presents a brief history of testing and assessment beginning with civil service testing in China in 2200 B.C. and ending with current debates on assessment. The authors try to point out how lessons from the past can help us in current endeavors.

Newmann, Fred. (February 1991). "Linking Restructuring to Authentic Student Achievement." *Phi Delta Kappan* 72, 6: 458-463. Makes the point that restructuring has to be goal-oriented. The first question to ask is "What changes do I want to see in students?" and then, "How will I restructure to achieve these student goals?"

Newmann, Fred. (1995). *A Guide to Authentic Instruction and Assessment: Vision, Standards and Scoring.* Madison: Wisconsin Center for Educational Research, University of Wisconsin, 1025 W. Johnson St., Room 2442, Madison, WI 53706. Phone: (608) 263-7575.

Newmann, Fred, and G. Wehlage. (1995). *Successful School Restructuring.* Madison: Wisconsin Center for Educational Research, University of Wisconsin, 1025 W. Johnson St., Room 2442, Madison, WI 53706. Phone: (608) 263-7575.

Paul, Richard W., and Gerald M. Nosich. (1992). "A Model for the National Assessment of Higher Order Thinking." In *Critical Thinking: What Every Person Needs to Survive in a Rapidly Changing World,* revised edition, edited by A.J.A. Binker. Available from: The Foundation for Critical Thinking, 4655 Sonoma Mountain Rd., Santa Rosa, CA 95404. Phone: (707) 546-4926. This paper provides a model for the national assessment of higher-order thinking. The first main section of the paper states 21 criteria for higher-order thinking assessment.

Paul, Richard W., and Gerald Nosich. (1992). "Using Intellectual Standards to Assess Student Reasoning." In *Critical Thinking: What Every Person Needs to Survive in a Rapidly Changing World,* revised edition, edited by A.J.A. Binker. Available from: The Foundation for Critical Thinking, 4655 Sonoma Mountain Rd., Santa Rosa, CA 95404. Phone: (707) 546-4926. Describes components of reasoning.

Stiggins, Richard J. (1994). *Student-Centered Classroom Assessment.* New York: Macmillan. A textbook on the use of different types of assessments (selected response, essay, performance, and personal communication) to evaluate students' attainment of knowledge, reasoning, skill, product, and affective outcomes. The focus is classroom assessment.

UCLA, Center for the Study of Evaluation and The National Center for Research on Evaluation, Standards, and Student Testing. (Winter 1994). *Evaluation Comment.* Los Angeles: UCLA, CSE/CRESST, Graduate School of Education. This issue deals mainly with the proceedings of the 1993 CRESST Conference, "Assessment Questions: Equity Answers."

UCLA, Center for the Student of Evaluation and The National Center for Research on Evaluation, Standards, and Student Testing. (Spring 1993). *Evaluation Comment,* Spring 1993. Los Angeles: UCLA, CSE/CRESST, Graduate School of Education. This issue deals mainly with the proceedings of the 1992 CRESST Conference, "What Works in Performance Assessment?"

**Literacy Assessment, Instruction, and Goals for Students:**

Calfee, Robert. (1988). "The Teacher's Role in Using Assessment to Improve Learning." In *Assessment in the Service of Learning.* Princeton, N.J.: Educational Testing Service. Describes current and past notions of reading instruction and how assessment relates to these notions.

Stayter, Francine, and Peter Johnston. (1990). "Evaluating the Teaching and Learning of Literacy." In *Reading and Writing Together: New Perspectives for the Classroom,* edited by T. Shanahan. Norwood, Mass. Christopher-Gordon. Describes how assessment, if done correctly, can be a tool for learning, not just a tool for monitoring.

**Mathematics Assessment, Instruction, and Goals for Students:**

Commission on Standards for School Mathematics. (1989). *Curriculum and Evaluation Standards for School Mathematics.* Reston, Va.: National Council of Teachers of Mathematics. Considered the bible for curriculum standards in mathematics.

Lacampagne, Carole B. (1993). *Transforming Ideas for Teaching and Learning Mathematics.* Washington, D.C.: U.S. Department of Education, 555 New Jersey Ave. N.W., Washington DC 20208. Phone: (202) 783-3238. An excellent overview of what is going on in mathematics goals for students, instruction, and assessment. A good publication for parents, school administrators, and others needing a quick overview.

Mathematical Sciences Education Board. (1993). *Measuring What Counts: A Policy Brief.* Washington, D.C.: National Academy Press. A good statement of the position that assessment should be a tool for instruction, not just for monitoring student progress.

Regional Educational Laboratories. (1993). *EdTalk: What We Know About Mathematics Teaching and Learning*. Washington, D.C.: Council for Educational Development and Research, 200 L Street N.W., Suite 601, Washington, DC 20036, (202) 223-1593. An excellent summary of current thinking about goals for students, instruction, assessment, and the role of the community.

Zarinnia, E. Ann, and Thomas Romberg. (1992). "A Framework for the California Assessment Program to Report Students' Achievement in Mathematics." In *Mathematics Assessment and Evaluation: Imperatives for Mathematics Educators*, edited by Thomas Romberg. Albany, N.Y.: State University of New York Press. This paper takes the position that assessment affects instruction and, therefore, regardless of the other purposes for the assessment, the instructional implications of our assessments must be taken into account.

### Science Assessment, Instruction, and Goals for Students:

McKinney, Kay. (1993). *Improving Math and Science Teaching*. Washington, D.C.: US Government Printing Office, Stock No. 065-000-00553-1, P.O. Box 371954, Pittsburgh, PA 15250-7954. Summarizes the presentations at the Secretary's Second Conference of Math and Science concerning the reasons science and mathematics instruction must change, the goals for such instruction, and the implications for materials, curriculum, assessment and teacher preparation.

National Research Council. (1994). National Science Education Standards—Draft. Washington, D.C.: National Committee on Science Education Standards, 2101 Constitution Ave., N.W. HA 486, Washington, D.C. 20418. Phone: (202) 334-1399.

Regional Educational Laboratories. (1993). *EdTalk: What We Know About Science Teaching and Learning*. Washington, D.C.: Council for Educational Development and Research, 2000 L Street N.W., Suite 601, Washington, DC 20036. Phone: (202) 223-1593. An excellent summary of current thinking about why science is important for all citizens, goals for students, instruction, assessment, what teachers need, and the role of parents.

### Writing Assessment, Instruction and Goals for Students:

Spandel, Vicki, and Richard Stiggins. (1989). *Creating Writers*. New York: Longman. How to improve student writing ability by teaching students to be self-assessors of writing.

### The Change Process:

Changing one's approach to assessment is the same as any other change: it takes time, resources, commitment, and training. These articles are included as good summaries of what it takes to make meaningful change in educational practice.

Fitzpatrick, Kathleen. (May 1992). "Restructuring to Achieve Outcomes of Significance for All Students." *Educational Leadership* 49, 8: 18-22. Describes one district's experience with restructuring.

Ledell, Marjorie, and Arleen Arnsparger. (1993). *How to Deal With Community Criticism of School Change*. Alexandria, Va.: ASCD. This thin publication has sections dealing with communication, building support, working with the media, and what to do when you're under attack.

Muncey, Donna, and Patrick McQuillan. (February 1993). "Preliminary Findings from a Five-Year Study of the Coalition of Essential Schools." *Phi Delta Kappan* 74, 6: 486-494. True restructuring is very hard, time-consuming, and takes training.

Phi Delta Kappan. (February 1993). "Is Your School Ready for Alternative Assessment?" *Phi Delta Kappan* 74, 6: 455-456. Lists ten conditions that are important indicators of a school's readiness to embark on any new alternative assessment effort.

Shoemaker, Betty. (1993). *Implementation of Oregon's Educational Act for the Twenty-first Century*. Eugene: Eugene School District, 200 N. Monroe St., Eugene, OR 97402-4295. An excellent summary of the elements important for successful change.

### Annotated Bibliographies

Arter, Judith A. (updated regularly). *Assessment Alternatives in Science*. Portland, Oreg.: Northwest Regional Educational Laboratory.

Arter, Judith A. (updated regularly). *Bibliography of Assessment Alternatives: Social Studies*. Portland, Oreg.: Northwest Regional Educational Laboratory.

Arter, Judith A. (updated regularly). *Math Assessment Alternatives*. Portland, Oreg.: Northwest Regional Educational Laboratory.

Arter, Judith A. (updated regularly). *Portfolio Resources*. Portland, Oreg.: Northwest Regional Educational Laboratory.

Arter, Judith A. (updated regularly). *Reading Assessment Alternatives*. Portland, Oreg.: Northwest Regional Educational Laboratory.

## Videotape Resources

*Alternatives for Measuring Performance*. IRI/Skylight Publishing, Inc., 200 E. Wood St., Suite 250, Palatine, IL 60067. Phone: 1-800-348-4474, ext. 220.

On November 21, 1991, the North Central Regional Educational Laboratory broadcast a national teleconference on assessment to sites throughout the United States. This videotape (60 minutes) of the teleconference profiles two schools using innovative assessments. A discussion by a distinguished panel of assessment experts follows. The videotape is accompanied by a 36-page guidebook that describes the characteristics of effective assessments, criteria for valid performance assessments, and recommended resources.

*Arts PROPEL Video Handbook*. Harvard Project Zero, Longfellow Hall, 13 Appian Way, Cambridge, MA 02138. Phone: 1-800-235-2132.

Arts PROPEL was initiated to develop nontraditional models of assessment appropriate for students engaged in the visual arts, music, and imaginative writing. This video (65 minutes) accompanies the Arts PROPEL Handbook for Visual Arts and provides an overview to the Arts PROPEL project. The program includes sections on PROPEL's guiding principles, students engaged in long-term visual arts projects, and the use of portfolios as vehicles for communication and reflection. Teacher testimonials lend authenticity and practicality to the project.

*Assessment and Technology*. The Center for Technology in Education, Bank Street College of Education, 610 W. 112th St., New York, NY 10025.

This video was produced in conjunction with a research project to determine how video and computer technology can best be used to assess learning in science and mathematics. The video begins with a description of one school's use of technology-based science assessments. The program then presents four examples of a performance assessment in which 11th graders work in teams to design and construct motorized devices to demonstrate their understanding of important concepts in applied physics. Printed scoring criteria are provided and viewers are encouraged to evaluate the student performances as they view the video.

*California Learning Record, The*. California Department of Education, Bureau of Publications, Sales Unit, PO Box 271, Sacramento, CA 95812. Phone: (916) 445-1260.

This video was produced by the California Department of Education to introduce the California Learning Record (CLR), a developmentally appropriate portfolio assessment process for young children. Filmed in schools with Chapter 1 programs in the San Diego Unified School District, the program shows parents, teachers, and students making contributions to the CLR.

*Classroom Assessment Workshops*. IOX, 5301 Beethoven St., Suite 109, Los Angeles, CA 90066. Phone: (310) 822-3275.

This is a series of 14 interactive videotapes dealing with classroom assessment. The tapes were developed by Northwest Regional Educational Laboratory and have running times from one hour to all day. The series includes these titles: *A Status Report on Classroom Assessment; Understanding the Meaning and Importance of Quality Classroom Assessment; Developing Assessments Based on Observation and Judgment; Paper and Pencil Test Development; Measuring Thinking in the Classroom; Assessing Reading Proficiency; Writing Assessment: Issues and Answers; Writing Assessment: Training in Analytical Scoring; Developing Sound Grading Practices; Understanding Standardized Tests; Assessing Mathematical Power; Assessment in the Science Classroom; Using Portfolios in Assessment and Instruction; and Facing the Challenges of a New Era of Educational Assessment.*

*"In the Classroom" Video Training Series*. IOX, 5301 Beethoven St., Suite 109, Los Angeles, CA 90066. Phone: (310) 822-3275.

This series includes "Putting Portfolios to Work" (1994) and "Writing from the Inside Out: Revising for Quality" (1994).

*Mathematics Assessment: Alternative Approaches*. National Council of Teachers of Mathematics, 1906 Association Dr., Reston, VA 22091. Phone: 1-800-235-7566.

This video (71 minutes) was produced in cooperation with South Carolina Television as a companion to the book *Mathematics Assessment: Myths, Models, Good Questions, and Practical Suggestions* by Jean Stenmark. Through classroom examples and expert commentary, the program explores teachers' implementations of a variety of assessment methods in their classrooms. A viewer's guide is included.

*Performance Assessment in the Classroom*. The Video Journal of Education, 549 W 3560 S, Salt-Lake City, UT 84115. Phone: 1-800-572-1153.

This program consists of two tapes: *Part 1: Performance-Based Assessment and Instruction* (29 minutes) and *Part 2: Creating Performance Tasks* (31 minutes). The tapes are designed to introduce teachers, administrators, and parents to the key principles of performance-based assessment and instruction. The program invites the viewer to consider ten key questions related to the design and use of classroom performance tasks. Numerous classroom examples and comments by teachers make this a practical and fast-paced resource for staff development.

*Portfolio Assessment in the Classroom.* The Video Journal of Education, 549 W. 3560 S., Salt Lake City, UT 84115. Phone: 1-800-572-1153.

This program consists of two tapes: *Program 1: Reflections of Learning* (29 minutes) and *Program 2: Utilizing Portfolios* (29 minutes). The tapes feature Bena Kallick, a noted expert on portfolio assessment. The program emphasizes the use of portfolios to document learning and stimulate self-evaluation by students. A variety of classroom examples illustrate the use of portfolio assessment in different subject areas and grade levels.

*Redesigning Assessment.* Association for Supervision and Curriculum Development, 1250 N. Pitt St., Alexandria, VA 22314. Phone: 1-800-933-2723.

This series contains three tapes: *Introduction* (24 minutes), *Performance Assessment* (32 minutes), and *Portfolios* (40 minutes). The tapes feature classroom illustrations of performance assessment in action with comments from students, teachers, administrators, and national experts. Each videotape includes a leader's guide offering suggestions for using the videos as part of a workshop for professional staff, parents, and boards of education. These tapes are available for rental or purchase.

*Standards NOT Standardization.* Center for Learning, Assessment, and School Structure, 39 Main St., Geneseo, NY 14454. Phone: (716) 243-5500.

This series contains four volumes: *Volume I—Re-thinking Assessment: Provocations and Issues; Volume II—Re-thinking Assessment: The District; Volume III—Re-thinking Assessment: The Classroom; and Volume IV—Re-thinking Assessment: The School.* The series is the centerpiece for a series of seminars on raising the standards of student performance through attention to authentic assessments. Each volume contains two videotapes and a comprehensive manual providing detailed suggestions for seminar leaders.

*Student Growth: Directions for Assessing and Communicating Learning.* Edmonton Public Schools, Center for Education, 1 Kingsway, Edmonton, Alberta, Canada T5H 4G9.

This video (16 minutes) describes the use of a portfolio system in the Edmonton Public Schools for assessing six essential learning outcomes identified by the district. The video shows examples of portfolio review conferences with staff, students, and parents.

*Teacher TV - Episode 11: Alternative Assessment.* NEA Professional Library, PO Box 509, West Haven, CT 06516. Phone: 1-800-229-4200.

Co-produced by the National Education Association and the Learning Channel, this video (22 minutes) contains vignettes of teachers using alternative methods for assessing student learning. One example describes the use of a computer and scanner for assembling a digital portfolio at the elementary level. A second vignette illustrates how a teacher involved her students in peer evaluation as part of a science fair project.

*Teaching and Testing in Maryland Today: Education for the 21st Century.* Maryland Assessment Consortium, 115 E. Church St., Frederick, MD 21701. Phone: (301) 694-1800.

This video (14 minutes) was developed for use with school staffs, parents, community members, and boards of education. It provides an overview of the new role of performance-based instruction and assessment in the context of Maryland's outcome-based educational reforms.

# X-3

## Contributors

**Dale Armstrong**, Director of Student Assessment, Centre for Education, Edmonton Public Schools, One Kingsway, Edmonton, Alberta, Canada T5H 4G9. Phone: (403) 429-8349; fax: (403) 425-3195; e-mail: darmstro@epsb.edmonton.ab.ca

**Dean G. Arrasmith**, Director, Evaluation and Assessment Program, Northwest Regional Educational Laboratory, 101 S.W. Main St., Suite 500, Portland, OR 97204. Phone: (503) 275-9557; e-mail: arrasmid@nwrel.org

**Judith A. Arter**, Unit Manager, Evaluation and Assessment Program, Northwest Regional Educational Laboratory, 101 S.W. Main St., Suite 500, Portland, OR 97204. Phone: (503) 275-9562

**Pamela R. Aschbacher**, Project Director, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Center for the Study of Evaluation, Graduate School of Education & Information Studies, UCLA, 10920 Wilshire Blvd, Suite 900, Los Angeles, CA 90024. Phone: (310) 206-1532

**Gail P. Baxter**, Researcher, Graduate School of Education, University of California, Santa Barbara, CA 93106-9490. Phone: (805) 961-3917

**Vern Begitschke**, Associate Principal, John Hersey High School, Township High School District #214, 1900 E. Thomas St., Arlington Heights, IL 60005. Phone: (708) 818-7608

**Paul Berman**, President, BW Associates, 815 Allston Way, Berkeley, CA 94710.

**Carol Black**, Education Specialist, School Reform, Oregon Department of Education, 700 Pringle Parkway S.E., Salem, OR 97310. Phone: (503) 378-8004

**Robert E. Blum**, Director, School, Community and Professional Development Program, Northwest Regional Educational Laboratory, 101 S.W. Main St., Suite 500, Portland, OR 97204. Phone: (503) 275-9615

**Linda Bond**, Director of Assessment, North Central Regional Educational Laboratory, 1900 Spring Rd., Suite 300, Oak Brook, IL 60521-1480. Phone: (708) 571-4700

**Jeanne Bond-Essner**, Lake Oswego School District, P.O. Box 70, Lake Oswego, OR 97034-0070. Phone: (503) 636-7691

**Vicki Boyd**, Research Associate, Services for School Improvement, Southwest Educational Development Laboratory, 211 East Seventh St., Austin, TX 78701-3281. Phone: (512) 476-6861

**Ernest L. Boyer** (deceased), President, The Carnegie Foundation for the Advancement of Teaching, 5 Ivy Lane, Princeton, New Jersey 08540. Phone: (609) 452-1780

**Jane B. Braunger**, Senior Research Associate, Literacy, Language and Communication Program, Northwest Regional Educational Laboratory 101 S.W. Main St., Suite 500, Portland, OR 97204. Phone: (503) 275-9588. e-mail: braungej@nwrel.org

**Kathleen U. Busick**, Program Specialist, Pacific Region Educational Laboratory, 828 Fort Street Mall, Suite 500, Honolulu, HI 96813. Phone: (808) 533-6000

**Jocelyn A. Butler**, Consultant, 3723 N.E. 21st Ave., Portland, OR 97212. Phone: (503) 284-8033

**Annie Calkins**, Director, Curriculum and Evaluation, Juneau School District 10014 Crazy Horse Dr., Juneau, AK 99801. Phone: (907) 463-1700

**Roberta Camp**, Independent Consultant, 713 S. Warnock St., Philadelphia, PA 19147. Phone: (215) 922-7907

**Carnegie Corporation of New York**, 437 Madison Ave., New York, NY 10022. Phone: (212) 371-3200

**Anthony P. Carnevale**, Vice President and Director of Human Resource Studies, Committee for Economic Development, 2000 L St. N.W., Suite 700, Washington, D.C. 20036. Phone: (202) 296-5860

**Kevin Castner**, Associate Superintendent of Curriculum and Instruction, Frederick County Public Schools, 115 East Church St., Frederick, MD 21701. Phone: (301) 694-1355

**Commission on Maine's Common Core of Learning**, Maine State Department of Educational and Cultural Services, State House Station 23, Augusta, MN 04333. Phone: (207) 289-5811

**David T. Conley**, Associate Professor, Division of Educational Policy and Management, College of Education, University of Oregon, Eugene, OR 97403-1215. Phone: (503) 346-5171

**H. Dickson Corbett**, Co-Director, Applied Research Program, Research for Better Schools 444 North Third St., Philadelphia, PA 19123. Phone: (215) 574-9300

**Arthur L. Costa**, Professor Emeritus, California State University, P.O. Box 2418, Granite Bay, CA 95746. Phone: (916) 797-1071 or 278-7737

**Lorraine Costella**, Assistant State Superintendent for Instruction, Frederick County Public Schools, Maryland Assessment Consortium, 115 East Church St., Frederick, MD 21701. Phone: (301) 694-1337

**Ruth Culham**, Research Associate, Evaluation and Assessment Program, Northwest Regional Educational Laboratory, 101 S.W. Main St., Suite 500, Portland, OR 97204. Phone: (503) 275-9583

**Linda Darling-Hammond**, Co-Director, National Center for Restructuring Education, Schools and Teaching, Box 110, Teachers College, Columbia University, 525 W. 120th St., New York, NY 10027. Phone: (212) 678-3432

**Donald Daws**, Mathematics Teacher, Metro High School, 1212 7th St, S.E., Cedar Rapids, IA 52503. Phone: (319) 398-2193.

**Terrence E. Deal**, Professor of Education, Box 514, Peabody College, Vanderbilt University, Nashville, TN 37203. Phone: (615) 322-8014

**Rodney Doran**, Professor of Science Education, Graduate School of Education, SUNY at Buffalo, Buffalo, NY 14260. Phone: (716) 645-3171

**Gary E. Dunkleberger**, Assistant Superintendent of Instruction, Carroll County Public Schools, 55 N. Court St., Westminster, MD 21157. Phone: (410) 751-3126; e-mail: garyed@umd5.umd.edu

**Education Daily**, Capitol Publications, Inc., P.O. Box 1453, Alexandria, VA 22313-2053. Phone: (703) 683-4100

**Elliot W. Eisner**, Professor, School of Education, Stanford University, Palo Alto CA 94305-2060. Phone: (415) 723-2300

**JoAnne Eresh**, School Support Specialist, The Connelly Center, 1501 Bedford Ave., Pittsburgh, PA 15219. Phone: (412) 338-8047

**Christine Sobray Evans**, Teacher, Bernardo Heights Middle School, 12990 Paseo Lucido, San Diego, CA 92128.

**Todd Fennimore**, Writer/Researcher, North Central Regional Educational Laboratory , 1900 Spring Rd., Suite 300, Oak Brook, IL 60521-1480. Phone: (708) 571-4700

**Steven Ferrara**, State Director of Student Assessment, Maryland State Department of Education, 200 W. Baltimore St., Baltimore, MD 21201. Phone: (410) 767-0081; e-mail: steve ferrara@umail.umd.edu

**Kathleen A. Fitzpatrick**, Executive Director, National Study of School Evaluation, 1699 East Woodfield Rd., Suite 406, Schaumburg, IL 60173. Phone: (708) 995-9080, (800) THE NSSE; fax: (708) 995-9088

**Michael G. Fullan**, Dean of Education, University of Toronto, 371 Bloor Street West, Toronto, Ontario, Canada M5S 2R7. Phone: (416) 978-3223

**Leila J. Gainer**, LHW&H, 3005 Highway 17 North Bypass, Myrtle Beach, South Carolina 29577. (803) 448-1123

**Gail Lynn Goldberg**, Specialist in Test Development and MSPAP Scoring Lead, Maryland State Department of Education, 200 W. Baltimore St., Baltimore, MD 21201-2595. Phone: (410) 767-0084; e-mail: gaillg@umd5.umd.edu

**Claude Goldenberg**, Associate Professor, Department of Teacher Education, California State University, 1250 Bellflower Blvd., Long Beach, CA 90840. Phone: (310) 985-4443; e-mail: cgolden@csulb.edu

**Thomas R. Guskey**, Professor, Educational Policy Studies and Evaluation, 131 Taylor Education Bldg., College of Education, University of Kentucky, Lexington, KY 40506. Phone: (606) 257-2813

**Gene V. Hall**, Dean, College of Education, University of Northern Colorado, Greeley, CO 80639. Phone: (303) 351-2861

**Greg Hall**, Assessment Specialist, Alberta Education, Box 43, 11160 Jasper Ave., Edmonton, Alberta, Canada TK5 OL2. Phone: (403) 427-0010

**Douglas E. Harris**, Director of Education, National Gardening Association, 180 Flynn Ave., Burlington, VT 05401. Phone: (802) 863-1308

**Joan Herman**, Associate Director, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Center for the Study of Evaluation, Graduate School of Education & Information Studies, UCLA, 10920 Wilshire Blvd, Suite 900, Los Angeles, CA 90024. Phone: (310) 206-1532; fax: (310) 825-3883

**Stephen K. Hess**, Director of Criterion-Referenced Evaluation and Testing, Frederick County Public Schools, 115 East Church St., Frederick, MD 21701. Phone: (301) 694-1355

**K. Michael Hibbard**, Assistant Superintendent, Region 15 Public Schools, P.O. Box 395, Middlebury, CT 06762. Phone: (203) 758-8250

**John Hillary**, Assistant Superintendent for Instruction, Township High School District 214, 2121 S. Goebbert Rd., Arlington Heights, IL 60005. Phone: (708) 364-8644; e-mail: jhill214@aol.com or jhillary @dist214.k12.il.us

**Shirley M. Hord**, Senior Research Associate, Southwest Educational Development Laboratory, 211 E. Seventh St., Austin, TX 78701-3281. Phone: (512) 476-6861; fax: (512) 476-2286; e-mail: shord@ sedl.org

**Peter Johnston**, Professor of Education, ED 330 Reading Dept., SUNY, 1400 Washington Ave., Albany, NY 12222. Phone: (518) 442-5104

**Beau Fly Jones**, Senior Researcher, North Central Regional Educational Laboratory, 1900 Spring Rd., Suite 300, Oak Brook, IL 60521-1480. Phone: (708) 571-4700

**Bena Kallick**, Educational Consultant, 12 Crooked Mile Rd., Westport, CT 06880. Phone: (203) 227-7261

**Alfie Kohn**, 242 School St., Belmont, MA 02178. Phone: (617) 489-6300; fax: (617) 489-6314.

**Rachel Korach**, Lake Oswego School District, P.O. Box 70, Lake Oswego, OR 97034-0700. Phone: (503) 636-7691

**Marilynn Kulieke**, Director of Research and Evaluation, Township High School District 214, 2121 S. Goebbert Rd., Arlington Heights, IL 60005. Phone: (708) 364-8642

**James W. Kushman**, Research Associate; School, Community and Professional Development Program, Northwest Regional Educational Laboratory, 101 S.W. Main St., Suite 500, Portland, OR 97204. Phone: (503) 275-9569; e-mail: kushmanj@nwrel.org

**William L. Lazzaro**, Classroom Teacher, Montgomery County Public Schools, 13710 Samhill Dr., Mount Airy, MD 21771. Phone: (301) 829-9607

**Paul G. LeMahieu**, Director, Delaware Education Research and Development Center, University of Delaware, College of Education, Newark, DE 19716-2976. Phone: (302) 831-4433

**Robert J. Marzano**, Director, Research and Development Program, Mid-continent Regional Educational Laboratory, 2550 South Parker Rd., Aurora, CO 80014. Phone: (803) 337-0990

**Terry Mazany**, Associate Superintendent for Curriculum and Instruction, Oakland Unified School District, 314 E. 10th St., Rm. 202, Oakland, CA 94606. Phone: 510-836-8253; e-mail: tmazany@ousd. k12.ca.us

**Joseph P. McDonald**, Senior Researcher, Coalition of Essential Schools, Brown University, Education Department, Box 1938, Providence, RI 02912. Phone: (401) 863-3384

**Jay McTighe**, Director, Maryland Assessment Consortium, c/o Frederick Public Schools, 115 East Church St., Frederick, MD 21701. Phone: (301) 694-1337; e-mail: jmctigh@aol.com

**Ann S. Meltzer**, Senior Analyst, KRA Corporation, 1010 Wayne Avenue, Suite 850, Silver Spring, MD 20910. Phone: (301) 495-1591

**Wayne M. Neuberger**, Assessment Supervisor, Oregon Department of Education, 255 Capitol Street N.E., Salem, OR 97310-0203. Phone: (503) 378-8004

**Fred M. Newmann**, Director, Wisconsin Center for Educational Research on Organization and Restructuring of Schools, 1025 West Johnson St., Madison, WI 53706. Phone: (608) 263-7557

**Pamala Noli**, Director, California School Leadership Academy, California Department of Education, 313 West Winton Ave., Suite 703, Hayward, CA 94544. Phone: (510) 887-8808; e-mail: sfpam@aol.com

**Nancey Olson**, Senior Associate; School, Community and Professional Development Program, Northwest Regional Educational Laboratory, 101 S.W. Main St., Suite 500, Portland, OR 97204. Phone: (503) 275-9617; e-mail: olsonn.nwrel.org

**Thomas A. Olson**, Senior Associate; School, Community, and Professional Development Program, Northwest Regional Educational Laboratory, 101 S.W. Main St., Suite 500, Portland, OR 97204. Phone: (503) 275-9629

**Jackie Palmer**, Senior Training /Technical Assistant Associate, Southwest Educational Development Laboratory, 11 East Seventh St., Austin, TX 78701. Phone: (512) 476-6861

**Fran Davis Perkins**, Assistant Professor, Department of Teacher Education, Box 870231, College of Education, University of Alabama, Tuscaloosa, AL 35487. Phone: (205) 348-2419

**Kent D. Peterson**, Director, National Center for Effective Schools, University of Wisconsin-Madison, 1025 W. Johnson St. Suite 685, Madison, WI 53706. Phone: (608) 263-4730

**Jim Pollard**, Northwest Regional Educational Laboratory, 101 S.W. Main St., Suite 500, Portland, OR 97204. Phone: (503) 275-9500

**Bill Saunders**, Associate Research Psychologist, Department of Psychiatry and Biobehavioral Sciences, 405 Hilgard Avenue, Los Angeles, CA 90024. E-mail: ieq2wxs@mvs.oac.ucla.edu

**Sandra Schnitzer**, Strategic Plan Facilitator, K-12, Instructional Services, Aurora Public Schools, 15751 E. First Ave., Aurora, CO 80011. Phone: (303) 340-0861

**Peter M. Senge**, Director of the Systems Thinking and Organizational Learning Program at MIT's Sloan School of Management, and a founding partner of Innovation Associates in Framingham, Massachusetts.

**Richard J. Shavelson**, Professor, Department of Education, University of California, Santa Barbara, Santa Barbara, CA 93106-9490. Phone: (805) 961-3917

**Lorrie A. Shepard**, Professor, School of Education, University of Colorado 3695 Folsom St., Boulder, CO 80304. Phone: (303) 492-8108

**Ron Smith**, Director, Curriculum and Management Information, Lake Oswego School District, P.O. Box 70, Lake Oswego, OR 97034-0070. Phone: (503) 635-0306

**Vicki Spandel**, Senior Associate, Evaluation and Assessment Program, Northwest Regional Educational Laboratory, 101 S.W. Main Street, Suite 500, Portland, OR 97204. Phone: (503) 275-9564; e-mail: spandelv@nwrel.org

**Doris Sperling**, Classroom Assessment Consultant, Ann Arbor, Michigan. Retired from the Ann Arbor Public Schools, 1265 Lincolnshire Lane, Ann Arbor, MI 48103.

**Francine Z. Stayter**, Assistant Professor, Department of Education, State University of New York, Oneonta, NY 13820. Phone: (607) 436-3891

**Richard J. Stiggins**, Director, Assessment Training Institute, 1206 S.W. Hume Ct., Portland, OR 97219. Phone: (503) 452-3816

**Joyce Stone**, Director of Student Support Services, Missisquoi Valley Union High School, RR2, Box 268, Swanton, VT 05488. Phone: (802) 868-7311

**Sam Stringfield**, Principal Research Scientist, Center for Research on Effective Schooling for Disadvantaged Students, Johns Hopkins University, 3505 North Charles St., Baltimore, MD 21218. Phone: (410) 516-0370

**Charles Teddlie**, Professor, Department of Education, Louisiana State University, Baton Rouge, LA 70803-4721. Phone: (504) 388-3202

**Barbara Benham Tye**, Associate Professor of Education, Chapman University (College), Orange, CA 92666. Phone: (714) 997-6815

**Louise F. Waynant**, Deputy Superintendent, Prince George's County Public Schools, 14201 School Lane, Upper Marlboro, MD 20772.

**Tim Westerberg**, Principal, Littleton High School, 199 East Littleton Blvd., Littleton, CO 80121. Phone: (303) 798-2584

**Grant Wiggins**, Director of Educational Research and Development, Center on Learning, Assessment, and School Structure (CLASS), 648 The Great Rd., Princeton, NJ 08540

**Bruce Wilson**, Co-Director, Applied Research Program, Research for Better Schools, 444 North Third St., Philadelphia, PA 19123. Phone: (215) 574-9300

**John Winbury**, Teacher/Consultant, Poway Unified School District, 14640 Pierra Bonita Rd., Poway, CA 92064.

**Peter Winograd**, Professor and Chair, Department of Curriculum and Instruction, College of Education, University of Kentucky, Lexington, KY 40506-0032. Phone: (606) 257-4661

**Lynn Winters**, Project Director, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Center for the Study of Evaluation, Graduate School of Education & Information Studies, UCLA 405 Hilgard Avenue, Los Angeles, CA 90024. Phone: (310) 206-1532

**Dennie Palmer Wolf**, Senior Research Associate, Harvard Graduate School of Education; Executive Director, Project PACE, Harvard University, Cambridge, MA 02138. Phone: (617) 496-2770

**Driek Zirinsky**, Professor of English, English Department, Boise State University, Boise, ID 83725. Phone: (208) 385-1822

**Samia Zumout**, California Learning Assessment System, California Department of Education, P.O. Box 944272, Sacramento, CA 94244. Phone: (916) 657-3011

Association for Supervision
and Curriculum Development
Alexandria, Virginia